Modelling Power Systems in a Long-Term Power Shortage

Alastair Heggie

Doctor of Philosophy

THE UNIVERSITY OF EDINBURGH

2019

Abstract

This thesis proposes and analyses decision support models for power systems operating in chronic power shortage conditions. Mostly these power systems exist in Sub-Saharan Africa, but other countries, for example Pakistan, Nepal, Cambodia and Bangladesh, suffer from similar problems. The thesis is structured in three parts looking at demand forecasting, distribution level load shedding and national level power rationing.

First, we develop methods to forecast the electrical load on a power system conditional on a policy of load shedding. Our methods are based on the Linear Gaussian State Space Model and the Kalman Filter. Conventional time series forecasting methods cannot be applied in a power system operating in a state of chronic load shedding because the observed demand is determined both by the latent unsuppressed demand and by the load shedding decisions of the Distribution System Operator. We demonstrate the accuracy of these forecasting methods on a dataset from a Nigerian electricity distribution company. In addition, these models have potential to improve estimates of the latent demand for electricity compared to existing methods that rely on unreliable proxy variables or 'bottom-up' calculations that are difficult to verify.

Next, we formulate an optimization problem to help Distribution System Operators incorporate probabilistic demand forecasts such as those developed in this thesis into their planning of load shedding. Our problem is closely related to a stochastic variant of the classic "knapsack problem" with random item "weights". We extend the literature on this problem to study the case where the item weights are given by a stochastic process which is only observed after an item is included in the knapsack. Our computational experiments provide evidence that the theoretical benefits of planning ahead are not realized in practice. It seems that for realistic range of stochastic demand processes a robust policy can be derived by an approach that only considers immediate costs and benefits.

Finally, we study the problem of balancing supply and demand at the national level. We develop an AC Optimal Power Flow model with endogenous load shedding and use this to quantify the trade-off between maximizing the total amount of power delivered and distributing the available power between regional distribution companies in a fairer way. Our model represents the situation in the Nigerian power system in which the system operator minimizes load shedding, subject to exogenous proportional power supply targets for different regions. We explore how the level of permitted deviation from the target and the time period over which this target is to be achieved affects the level of load shedding. The use of an AC power flow model complicates the problem but is necessary because voltage constraints are often binding in highly stressed transmission networks in countries like Nigeria.

Acknowledgements

I would like to thank my supervisors, Dr Harry van der Weijde and Professor Ken McKinnon for their guidance and encouragement at every stage of my doctoral studies. In addition Dr Amy Wilson provided some insightful comments on an early version of Chapter 3.

The research documented in this thesis would not have been possible without cooperation and financial support from Energy Market and Regulatory Consultants (EMRC). I am particularly grateful to Alice Waltham and Dan Eager for their advice throughout.

I am grateful for the love and support of my mother and father, Bill and Hilary Heggie.

Lastly, I would like to thank Laura Harrison whose love, understanding and patience has been invaluable to me.

The value of this thesis is far greater for the help of all those mentioned above, while its faults are mine alone.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Alastair Heggie

Contents

Figures and Tables

Figures

Tables

Chapter 1

Introduction

In 2008 Ebhard et al. wrote, ‘Sub-Saharan Africa is in the midst of a power crisis marked by insufficient generating capacity, unreliable supplies, high prices, and low rates of popular access to the electricity grid’ [18]. 10 years have passed, and this remains a broadly accurate description for Sub-Saharan Africa and indeed for many other developing countries.

Power crises in Less Economically Developed Countries (LEDCs) manifest themselves in three key ways. Firstly, we can look at power crises from the point of view of access to electricity. Just 30% of the population in Sub-Saharan Africa is connected to the grid and 60% in South East Asia [33] [1]. Secondly, for those who are fortunate enough to have a grid connected electricity supply, the connection is often highly unreliable. Grid power supply is characterized by large voltage fluctuations and is frequently interrupted. In many countries utilities are forced to impose rolling blackouts to balance supply and demand. Finally, we can take the financial perspective. Paradoxically, average tariffs in many LEDCs are high by global standards and yet utilities are not able to cover their costs. This is a result of rampant power theft and non-payment combined with failures to procure generation forcing grids to rely on expensive emergency power contracts.

The literature on how to reform power systems in these conditions is extensive yet fraught with controversy. We do not intend to engage in this debate. We take the view that Eberhard et al. are correct to write, “The prerequisite for solid sector financing is better operating performance and thus greater financial viability by the incumbent utilities”[18]. Therefore, we will be concerned with how existing infrastructure and institutions could be better operated to improve welfare. To this end this thesis proposes and analyses decision support models for power systems operating in chronic power shortage conditions. The use of decision support models and optimization methods is routine in the power sector to solve problems of economic dispatch, hydro/thermal scheduling, Optimal Power Flow (OPF), fuel scheduling and demand forecasting. We seek to adapt models of this general application area to the unique challenges faced in a power crisis.

1. Electricity access in general is a complex and multi-faceted concept that cannot be measured simply by access to grid electricity. Instead of binary measures of access it is increasingly common to speak of multi-tier energy access levels which consider reliability of supply and off grid forms of access such as solar home systems. Despite this caveat, increasing the reach and reliability of electricity grids is an important part of the puzzle of increasing electricity access in LEDCs.

The scope of this thesis is therefore all power systems meeting the challenges identified above. Mostly these power systems exist in Sub-Saharan Africa, but other countries, for example Pakistan, Nepal, Cambodia and Bangladesh, suffer from similar problems. We will pay particular attention to the case of Nigeria. This is motivated in part by practical considerations; through contacts in consulting firms we have been able to obtain access to data on Nigeria. We will attempt to draw lessons from the Nigerian case that can be applied to all power systems in a similar situation.

We study 3 aspects of operational modelling of power systems. We now discuss each in turn.

1.1 Part 1: Demand Forecasting

In a chronic power shortage demand cannot be directly measured from observed load because this does not take into account the unsatisfied demand due to load-shedding. We develop two methods for modelling electricity demand using the Linear Guassian State Space Model (LGSSM). Using these we can forecast the latent demand without load shedding and the electrical load conditional on a particular load shedding policy. We also consider how our demand forecasting approaches can be applied to give a post-hoc reconstruction of the latent electricity demand that is reduced by load shedding. This is useful for constructing electricity demand scenarios which can inform network planning and operation.

Our first method models the demand on individual distribution feeders. We form an explicit model for the stochastic and periodic components of demand on each feeder by representing it with a *Structural Model*. We assume that in periods of load shedding the demand process on each feeder is simply “hidden” from view and continues to evolve according to the same process. Under this assumption the problem is therefore one of modelling time series in the presence of a large amount of missing data. The state space approach to time series analysis provides a principled and effective means of handling missing data because the Kalman Filter algorithm estimates the most likely state of the system during periods of missing data conditional on the observed data. We refer to this as the Agreggated Structural Models (ASM) method. The ASM method can be extended to model behavioral responses to load shedding by including exogenous regression variables that measure things like, whether the feeder was disconnected in the previous hour, or how long the feeder has been disconnected for.

We propose an alternative to this method based on a Dynamic Regression Model (DRM) in which demand is expressed as a linear function of the number of connected customers with time varying regression coefficients. We show that, if customers are disaggregated into a small number of categories that are roughly homogeneous, (for example residential, large industrial or small industrial) regression coefficients that correspond to the latent demand from each category of customer can be inferred as the states of a linear Gaussian state space model.

Time varying coefficient regression models are common in the econometrics literature, but the novelty of our approach is to use the dynamics of the structural model to capture stochastic and periodic variation in the regression coefficients. We refer to this as the DSR approach.

We test our methods using a new dataset obtained from a Nigerian electricity distribution company and demonstrate that the ASM approach provides better forecasts than the DSR method. We go on to show that the best forecasting performance can be obtained by combining the two models by incorporating the output of the DSR model as an exogenous regression variable in the ASM model.

Our work builds upon the state space frame work for time series analysis developed by [26] and [17]. To our knowledge no authors have considered models for load forecasting during periods of persistent load shedding. Structural modelling approaches have been applied to electricity demand forecasting for example in [52] and [51]. These references use the "innovations" state space model which does not require the Kalman filter, but we prefer the "multiple source of error" formulation because this allows missing values to be handled in a principled way by the Kalman filter [5]. Time varying parameter models in which unobserved regression coefficients evolve according to a random walk or stationary process have been studied in the econometrics literature (for example in [12] and [24]). By using a structural model for the coefficients our model is better able to handle periodic variation in the coefficients.

1.2 Part 2: Distribution Level Load Shedding

The System Operator (SO) can only control loads directly by disconnecting large areas at the interface between the transmission and distribution systems. Therefore, the responsibility for routine load shedding to maintain the generation/load balance falls to Distribution System Operators (DSOs). DSOs can control the power drawn by connecting or disconnecting distribution feeders. The role of the SO is to instruct the DSOs how much electricity they should consume.

A variety of considerations affect the DSOs decisions. These may include: maximizing utilization of available energy; avoiding overconsumption of energy relative to the SO's instructions; providing a predictable, if intermittent, power supply to customers; meeting service quality standards such as minimum connection frequency or maximum outage length; and maximizing revenue by directing energy to feeders with lower collection losses and higher average tariffs. This is a difficult problem because the underlying demand process is stochastic and is also imperfectly observable because it is "hidden" by regular disconnection of feeders.

We formulate an optimization problem to help DSOs incorporate probabilistic demand forecasts such as those developed in this thesis into their demand management processes. Our problem is closely related to a stochastic variant of the classic "knapsack problem" with random item "weights". In the *Feeder Scheduling Problem*, the volume of the knapsack can be interpreted as the power allocated to a DSO and the randomly distributed item weights are the power demand on each feeder. The existing literature on this topic (see [40] and [35] for example) can be used to optimize a simplified case where the problem is solved for a single period and demand forecast errors are assumed to be uncorrelated. We extend this literature to study the case where the item weights are given by a stochastic process which is only observed after an item is included in the knapsack.

We give an Mixed Integer Non-Linear Program (MINLP) formulation of this problem and suggest a linear approximation that is more tractable. Our formulation is flexible enough to be extended in several ways. We can consider the case where the stochastic processes determining the items weights are correlated. We can aim to provide a more predictable power supply by minimizing the deviation of the policy from a predefined schedule. We can incorporate constraints that enforce a certain minimum supply quality for all feeders.

1.3 Part 3: National Level Load Shedding

In the final section we consider load shedding from a national perspective. We develop an AC OPF model with endogenous load shedding and use this to quantify the trade-off between maximizing the total amount of power delivered and distributing the available power between regional distribution companies in a fairer way. Our model represents the situation in the Nigerian power system in which the system operator minimizes load shedding, subject to exogenous proportional power supply targets for different regions. We explore how the level of permitted deviation from the target and the time period over which this target is to be achieved affects the level of load shedding. The use of an AC power flow model complicates the problem but is necessary because voltage constraints are often binding in the highly stressed transmission networks in countries like Nigeria.

To apply the model over longer time scales we need to represent changing levels of demand and generation availability by a set of operating points resulting in a large set of OPF problems linked by common constraints defining the proportional load allocation targets. The resulting minimization problem is mathematically challenging, but we develop a solution approach based on Lagrangian decomposition which is solved efficiently. These methods would also be applicable to several other regulatory approaches to managing the trade-off between efficiency and regional fairness. In addition, they can be used to evaluate network expansion projects. We demonstrate this by analysing the benefit of deploying reactive support in the Nigerian Power System.

Sophisticated models have been proposed for efficient allocation of service interruptions in power shortage conditions in [39], [42], [16] and[7]. However, these rely on complex financial mechanisms which are unlikely to be practical in developing countries where institutional competence is often poor [1]. There is literature on optimal load shedding such as [21], [14] and [62]. However, these treat load shedding as an emergency measure to return the system to a steady state in the case of a contingency event. The contribution of our work is to consider the implications of the case when load shedding is a constant necessity for the operation of the system. Our approach builds on the well-studied AC OPF model to provide useful decision support to system operators in highly stressed power systems. The material in this chapter is based on our previously published work [27].

1.4 Thesis Structure

In Chapter 2 we give an overview of the problems faced by power systems in many LEDCs. Section 2.1 summarizes the challenges under three broad headings. Section 2.2 gives a historical overview of the trends in developing world power systems and how they relate to the ongoing power crisis. Lastly, in Section 2.4 we elaborate on these issues in the specific case of Nigeria which will be used as an illustrative case study throughout the thesis.

Chapter 3 describes our work on demand estimation. Section 3.2 introduces the theory of the LGSSM and the Kalman filter which we use throughout this chapter. Section 3.3 develops two approaches to forecasting suppressed electricity demand based on the LGSSM. Section 3.4 compares the performance of these methods when applied to data from a Nigerian utility company. We demonstrate how the methods for forecasting suppressed demand can be used to estimate the latent unsuppressed demand in Section 3.5.

Chapter 4 shows how the forecasting models of the previous chapter can be used to plan distribution level load shedding. In Section 4.1 we identify the considerations that DSOs may take into account when determining load shedding schedules. In Section 4.2 we introduce the stochastic knapsack problem with random weights and in Section 4.3 we go on to show that the considerations affecting the DSO's load shedding decisions can be modelled in this framework. We give a detailed model formulation in Section 4.4. Section 4.6 describes some numerical experiments that were carried out to evaluate the model.

Chapter 5 describes our work on national level power rationing. In Section 5.2 we formulate an OPF model with dispatchable loads and constraints that model the requirement for different regions of the network to receive a given proportion of total load supplied. In the same section we show how these constraints can be enforced over different time horizons and develop solution methods. Section 5.3 applies the methods to the Nigerian case.

1.5 Contributions of this Thesis

This thesis makes some contributions that will be of interest in the context of power systems operating in a chronic power shortage. Firstly, we develop accurate short-term forecasting methods for electricity demand in chronic power shortage conditions. These methods also have applications for post-hoc reconstruction of the unsuppressed electricity demand. Secondly, we formulate a decision support model to help DSOs plan load shedding using the multi period stochastic knapsack problem. Thirdly we develop methods that the SO or regulator can use in a power shortage to quantify the trade-off between maximizing the total amount of power delivered and distributing the available power in a fairer way. We quantify the trade-off between regional equity and total power supply in the specific case of Nigeria, showing that current Nigerian policies reduce the total amount of power delivered by up to 5%.

We also make some general methodological contributions which will be of interest outside the field of energy system modelling. We show how to extend time varying coefficient regression models to the case of periodic variation in the coefficients modelled by trigonometric terms. We extend the stochastic knapsack problem with random weights to a multi-period case where the random weights are the output of a stochastic process and the forecast variance is endogenous to the model. Although this problem is difficult to solve in general, we give a tractable formulation in the case where the stochastic variation in the item weights are the output of an autoregressive process.

Chapter 2

Context

2.1 Power Crises in the Developing World

2.1.1 Low Rates of Electricity Access

Just 30% of the population in Sub-Saharan Africa and 60% in South East Asia are connected to the electricity grid [33]. Grid based power allows for economies of scale in providing generation capacity, so a grid-based solution potentially offers the least cost way to expand electricity access where population density is high enough. However, utilities in many Less Economically Developed Countries (LEDCs) struggle to cover their operational costs, let alone finance grid expansion, so a large population who could be economically served by the grid are left without a connection. As trends towards urbanization continue this is only likely to increase. Measured more broadly than grid connections, World Bank data from 2016 indicates that the rate of electricity access was 43% in Sub-Saharan Africa (excluding high income countries) and 86% in South Asia. By comparison, electricity access is nearly universal in all other regions.

While these statistics above are correct in broad terms - access to electricity is lowest in South Asia and Sub-Saharan Africa - they only measure access to electricity in a binary way based on yes-no indicators like “having a household electrical connection” and “using electricity for lighting” [58]. This obscures differences in electricity access in terms of quality, reliability and affordability and differences in energy access for personal use, industry and community facilities. To address this the Multi-Tier Energy Access Framework [19] measures energy access on dimensions of peak capacity, duration of availability, reliability, power quality, affordability, legality and health and safety.

Recognizing the multi-tiered nature of energy access makes it clear that many people can get better access to modern energy services through mini-grids, solar home systems, private diesel generators or other very small-scale electrification projects. These types of projects can play and important role in improving electricity access if large capital investments cannot be financed and for those people in geographic areas where grid expansion is not feasible. Nevertheless, for many people expanding the national grid will provide the least cost means of electricity access. Therefore, strengthening electricity utilities is important.

2.1.2 Poor Quality Power Supply

Regular blackouts are the norm across much of Sub-Saharan African and many other LEDCs. Power shortages can be caused by failure to procure and maintain enough transmission and generation capacity to keep pace with expanding populations and economies. They can also be caused by shortages of primary energy resources like coal and gas or inadequate infrastructure to supply them to generators. In some cases, blackouts are caused by component failure as a result of poor maintenance.

Unplanned-capacity loss factors and measures of the length and frequency of interruptions in transmission and distribution are the usual measures of power system reliability. However, these figures are not collected or reported reliably by most Sub-Saharan African countries. An alternative measure of power system reliability comes from the World Bank enterprise surveys which ask businesses questions relating to the frequency and duration of electrical power outages[54]. Table 2.1, reproduced from [20], summarizes the World Bank data for a selection of countries. In the most recent year available businesses reported 33 outages per month in Nigeria (2014), 28 per month in Benin (2016), 22 per month in Niger (2017), 21 per month in The Gambia (2018) and 17 per month in Burundi (2014). The severity of the problem is reflected in the pervasiveness of backup diesel generations. Also included in Table 2.1 is the quantity of backup generator capacity as a percentage of grid capacity estimated by [20] and the minimum and maximum of the distribution in parenthesis. This is estimated to be as high as 46% in the Democratic Republic of the Congo, 22% in Nigeria and 20% in Niger.

Table 2.1: Measures of power system reliability in Sub-Saharan Africa

Country	Number of electrical outages in a typical month	Duration of a typical electrical outage (hours)	Average annual outage hours	Backup generator availability (as percentage of grid capacity)
Angola	4.7	13.5	760	8% (1%–25%)
Cameroon	7.6	8.7	790	1% (1%–51%)
Cote d'Ivoire	3.5	5.5	230	6% (1%–22%)
Democratic Republic of the Congo (DRC)	12.3	5.6	830	46% (1%–51%)
Ethiopia	8.2	5.8	570	1% (1%–12%)
Ghana	8.4	7.8	790	12% (1%–22%)
Kenya	6.3	5.6	420	7% (1%–12%)
Mozambique	1.6	4.3	80	1% (1%–25%)
Niger	22	5.2	1400	20% (1%–22%)

Country	Number of electrical outages in a typical month	Duration of a typical electrical outage (hours)	Average annual outage hours	Backup generator availability (as percentage of grid capacity)
Nigeria	32.8	11.6	4600	22% (1%–22%)
Senegal	6	1.8	130	1% (1%–25%)
South Africa	0.9	4.5	50	2.5% (1%–25%)
Tanzania	8.9	6.3	670	12% (1%–12%)
Zambia	5.2	2.8	180	3% (1%–25%)
Zimbabwe	4.5	5.2	280	5% (1%–25%)

2.1.3 Financial Unsustainability

Even though electricity prices in many Sub-Saharan African countries are high by global standards, in no Sub-Saharan African countries do customers pay prices that allow for full cost-recovery [18]. Expensive generation, high technical and non-technical losses and a legacy of poorly targeted subsidies have created chronically under financed power systems.

A persistent cause of financial difficulties for utilities in Sub-Saharan Africa is high rates of non-technical losses. From the perspective of large-scale electrical power systems, electrical losses are the difference between energy generated and energy sold to consumers. Technical losses are the proportion of the loss that is accounted for by power dissipation in components of the transmission and distribution system. The remainder of the loss is called non-technical losses. Non-technical losses are not intrinsic to the physical power system but are the result of failure to accurately meter, bill and collect payment from energy consumers. Non-technical losses come from many sources. Customers with a legitimate connection can commit power theft by meter bypassing or tampering. In other cases, illegitimate connections are made to the power grid. Customers might also bribe or threaten utility staff to avoid payment. Some utilities are simply badly administered and fail to issue bills and enforce payment. If there is a weak regulatory regime or a corrupt legal system utilities may have little recourse if customers refuse to pay their bills.

[49] observes that features common to countries with high non-technical losses are poverty and prolonged economic, social and political turmoil: “In tumultuous times government organizations cease to function efficiently, become prone to corrupt practices, investment is not made in system management, and the consumers take advantage of the system”. Low willingness to pay can become entrenched if service quality is poor. The result is a vicious cycle of underinvestment and declining service quality further eroding willingness to pay.

2.1.4 Scope of the Problems

The pervasiveness of these problems can be seen by studying the UN Sustainable Energy For All (SE4ALL) initiative. SE4ALL is a scheme for accessing donor funding in pursuit of UN sustainable development goal 7 (ensuring universal access to modern energy services; doubling the rate of improvement in energy efficiency; and doubling the share of renewable energy in the global energy mix). As a condition of participating in the initiative and accessing donor funds countries were required to submit Rapid Action and Gap Analysis reports assessing the state of their energy system. Of 24 RAGAs submitted by Sub-Saharan African countries we have analysed the 14 submitted in English. All of these identified low rates of access to electricity, poor quality gird power supply and financial unsustainability of the power sector as problems. The same themes can be identified in RAGAs submitted by several South Asian countries: Pakistan, Cambodia, Nepal and Bangladesh. Table 7.1 in the appendix summarizes the common narratives from the English language RAGA documents submitted by countries in Asia and Sub Saharan Africa.

2.2 Historical Background

2.2.1 Early Expansion of Electricity Sector in Sub-Saharan Africa

In the post-colonial era development of electricity systems in Africa was driven largely by a view of electricity as a public good for the promotion of economic and social development [47]. Large infrastructure projects like hydroelectric dams and the provision of subsidized power to industry were seen as ways to promote development and modernization. Vertically integrated, often state-owned utilities were the norm. Although universal access was not a general goal, these policies were often successful at expanding access to electricity.

However, financial crises that spread across may African nations in the 1980s resulted in years of stagnation. Capacity was not expanded and fell into disrepair. Financial and technical management of utilities was poor. The legacy of policies of highly subsidized power meant that the power sector could not cover its costs. A minority enjoyed subsidized access to power while a majority had no access at all.

2.2.2 The 'Standard Prescription' is Applied to the Developing World with Mixed Results

Based on the early experience of the UK, Chile, Norway and the USA a template for electricity sector reform in developing countries emerged that has come to be called the "Standard Prescription" [30]. This model was promoted by Development Finance Institutions including the World Bank, Asian Development Bank, Inter-American Development Bank and UK

Department for International Development as well as a collection of consultants involved in early reforms [22]. The motivating situation and objectives of reform in developing economies was fundamentally different to that in the developed world. In developed countries the aim of reform was to bring down consumer prices by encouraging greater efficiency from well-functioning industries that already provided near universal access to reliable power. In the developed world the aim was to address poor financial and technical management of utilities, reduce state subsidies, bring about cost recovery through reductions in non-technical losses and tariff increases, and encourage private investment to grow the industry and increase energy access. The resulting policy prescription described in [22], [61] and [30] can be summarized by these steps: Corporatization to separate the utility from government departments; Commercialization to bring about cost recovery in pricing and enforcement of collections; new energy legislation to provide a basis for restructuring and private/foreign participation; establishment of a regulator independent from government; new power supply to be procured from Independent Power Projects (IPPs) with long term Power Purchase Agreements; Vertical and horizontal unbundling; Privatization of the unbundled generation and distribution companies; establishment of competitive wholesale and retail markets.

2.2.3 The Emergence of Hybrid Electricity Markets

While many developing countries have attempted some form of reform of their electricity industries, they have rarely proceeded in the sequence suggested by the standard prescription [22] and most are very far from bid-based competitive pools; highly managed markets with single buyers are the norm [61]. Evidence of the benefits of privatization in developing countries is unclear. A quantitative analysis of private sector participation in water and electricity utilities [56] concludes that private sector delivers higher labour productivity and operational efficiency, but no effect on investment levels or prices. The study was unable to say whether these efficiency gains are captured as higher profits or likely to reduce the burden on state finances. Furthermore, attracting Foreign Direct Investment depends in large part on the uncontrollable global investment climate [61].

Most Sub-Saharan African countries are now characterized by what [18] has called *hybrid markets*. These are power markets with elements of private sector participation such as IPPs and partially unbundled and corporatized utilities, but in which a state-owned utility retains a dominant position.

2.3 Improving the Performance of African Utilities

It is not the intention of this thesis to analyze how institutional and regulatory change could set developing world power systems on a better course. Nevertheless, for context we note that developing countries can look to examples where loss making utilities placing an unsustainable burden on national finances have been transformed. Examples from South America and India are particularly pertinent, showing how utilities can dramatically reduce non-technical losses. Several lessons can be drawn from the examples given in [57] which we briefly summarize here. A more detailed summary of each case is given in Appendix 8.

Firstly, effective loss reduction is a partnership between utilities, law enforcement, government and regulators; it relies upon credibly demonstrating government's commitment to tackling theft and creating a norm against electricity theft by publicizing cases of electricity theft and enforcement. Secondly, effective technical measures such as tamper proof and prepayment meters or medium voltage distribution networks can be highly effective but are expensive; They should therefore be targeted based on careful analysis to identify where they will have maximum impact. This often means that large consumers, who may have substantial political power, must be targeted. This reinforces the need for government and law enforcement to be committed partners in loss reduction. Furthermore, the analysis required to target loss reduction efforts effectively requires that business practices be re-engineered around ensuring reliable customer databases so that thieves can be detected efficiently through risk profiling.

Lastly, the link between tariffs, loss reduction and demand should not be ignored. Experience in the cases analysed in [57] shows that making users pay for power that previously was not billed/collected results in significantly reduced demand. Nevertheless, "In all successful cases, a large share of non-technical losses was concentrated in users able to pay for cost-reflective tariffs". Toleration of losses is an implicit subsidy for those consumers who do not pay by those who do. It may still be good policy to subsidize some of these consumers for reasons of social policy. The experience of Argentina illustrates that redistributive aims in the electricity supply industry can be achieved through explicit subsidies with an honest accounting of costs. By allowing distribution companies to recover their full costs from consumers, they have appropriate incentives expand electricity access. Serious consideration should be given to how best to explicitly subsidize customers who cannot afford to pay a cost reflective tariff in a system which generates enough revenue to do so.

2.4 The Nigerian Case Study

2.4.1 History

Nigeria's power system was developed under a state-owned system that began with the establishment of the Electricity Corporation of Nigeria 1951. This system was successful at expanding capacity until investment stagnated in 80s and 90s [4]. By the early 2000s the electricity supply industry had suffered decades of underinvestment and was hamstrung by poor revenue collection and tariffs that did not reflect its costs. The result was poor service quality with constant load shedding and regular system collapse. This energy crisis as supply failed to keep up with demand prompted the government of Nigeria to begin a program of power sector reform and privatization.

In 2001 Nigeria announced the intention to restructure the electricity supply industry and privatize the National Electric Power Authority (NEPA). Privatization of NEPA's assets was completed in November 2013, signalling the commencement of the first stage of the reformed industry.

2.4.2 Industry Structure

The ultimate destination of Nigeria's energy market reforms is not clear. At present the system is an example of a hybrid power market in which extensive state intervention in the power market coexists with private investment in IPPs and electricity distribution companies.

The key market participants are 11 privately owned, regional monopoly distribution and retail companies (referred to as Distribution Companies (DisCos) in Nigeria), numerous Generation Companies (GenCos), a government backed wholesale trader (Nigerian Bulk Electricity Trader (NBET)) and the state owned Transmission Company of Nigeria (TCN) which also has a system/market operator function. Electricity is traded by power purchase agreements which DisCos and GenCos enter into with the wholesale trader. The Nigerian Electricity Regulatory Commission (NERC) sets distribution and generation tariffs through the Multi-Year Tariff Order (MYTO) regulatory framework. The dispatch of the power system is centrally managed rather than the result of market mechanism. The relationships between the current market participants are summarized in Figure 2.1.

Below we compare the structure of the Nigerian electricity industry today to the Standard Prescription for electricity market reform.

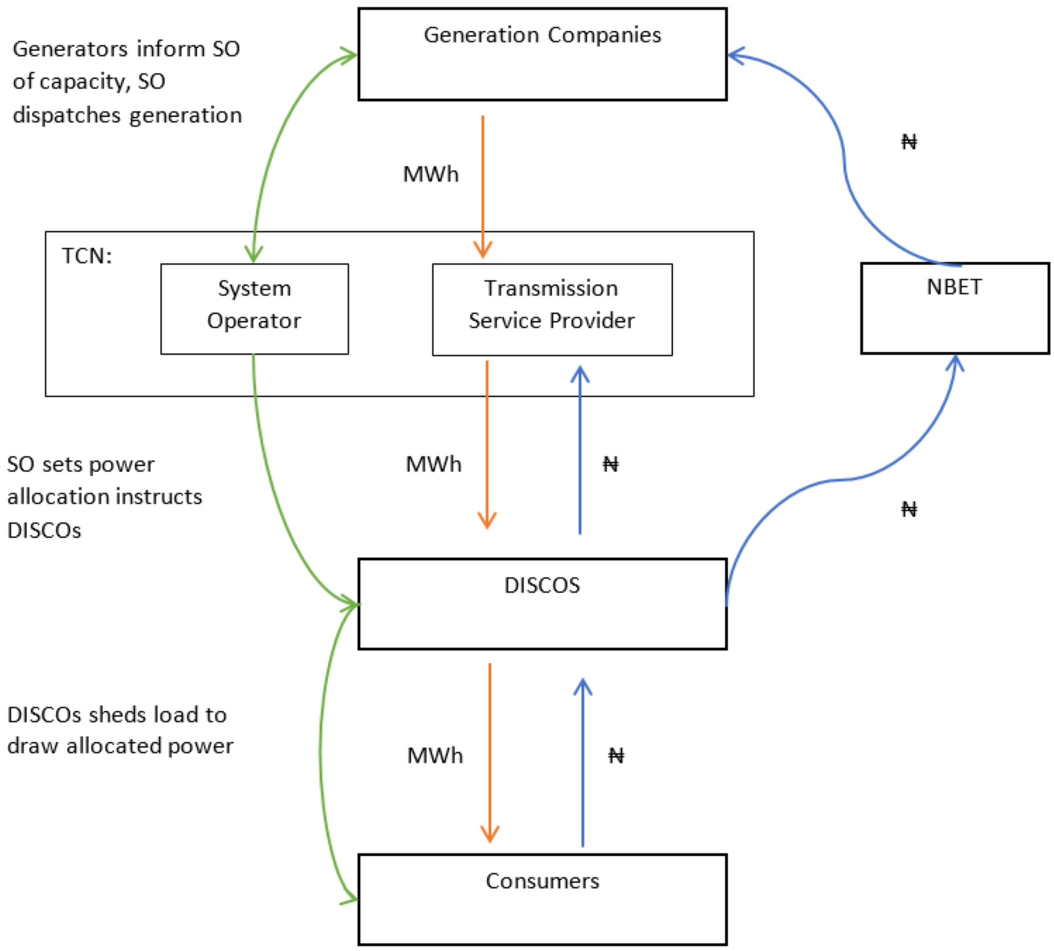


Figure 2.1: Structure of the Nigerian electricity supply industry

Fully implemented

- Enacting legislation to allow restructuring and creation of regulatory agencies - Electric Power Sector Reform Act (EPSRA) in 2005, inauguration of NERC October 2005
- Corporatization and commercialization of state-owned utilities - Nigeria Electricity Supply Company unbundled into 18 companies under flag of Power holding Company of Nigeria: 11 distribution companies, 6 generation companies and 1 transmission company and system/market operator
- Designation of an independent system operator – Fulfilled by TCN
- Privatization – Privatization of successor GenCos and DisCos completed November 2013. Niger Delta Power Holding Corporation to privatise remaining generation assets built under Nigerian National Integrated Power Project scheme in 2000s. Private management installed in TCN under contract with Manitoba Hydro

Partially implemented

- Separating potentially competitive elements from natural monopoly elements – Generation and distribution companies separate from transmission company, DisCos retain regional monopolies
- Independent power producers – IPPs predate EPSRA, but NERC regulation makes it difficult to enter the market on the grid unless NBET is undertaking bidding for capacity. Furthermore, third party generation is illegal, so entering into a private contract with a large customer who is willing to pay, is not possible

Not implemented

- Regulating to promote efficient access transmission network
- Creation of wholesale markets for energy and ancillary services
- Separation of competitive elements of retail tariffs from regulated network charges

2.4.3 Electricity Supply Quality

The Nigerian Power system fails to meet total electricity demand. As a result, power shortages have to be managed though load shedding at all hours of the day. The System Operator (SO)'s direct control of load is limited to disconnecting large areas at the interface between the transmission and distribution systems. Therefore, most load shedding is undertaken by the DisCos. The DisCos can control the power drawn by connecting or disconnecting distribution feeders. The role of the system operator is to give instructions to the DisCos telling them how much electricity they should consume.

A core part of NERC's policy, imposed under the MYTO, are proportional Load Allocation targets for each DisCo region. The precise implementation of this policy has varied since the introduction of the MYTO, but the principle is that the level of load supplied in each DisCo region over some time period should be a fixed proportion of the total level of load supplied. The Load Allocation roughly corresponds to the proportion of customers in each DisCo. TCN sets the daily Load Allocation which instructs DisCos how much power they are to consume. This allocation may be revised in response to actual system conditions. The resultant load is highly erratic because: a) the DisCos have poor knowledge of what the likely load is from any given feeder; and b) poor maintenance means that circuit components regularly fail or trip.

2.4.4 Electricity Network

Approximately 80% of the on-grid generation capacity of the Nigerian system is supplied by gas fired power stations. These are all located in the south of the country. The remainder is supplied by 3 hydro-electric power stations located on the Niger River in the central western part of the country. Loads are more evenly distributed over the country than generation. As such the transmission system has to carry a lot of power over large distances to the north.

TCN produces daily reports detailing the daily operations of the power system. These show that there is massive underutilization of capacity. The theoretical maximum capacity at all power generating sites connected to the national grid is 11,345 MW. However, in the period from privatization in November 2013 to January 2016 the average available generation capacity was 4281MW.

Despite having substantial reserves of natural gas, Nigeria lacks adequate infrastructure (in terms of reliability and capacity) to supply its thermal generators with fuel. Most gas extracted in Nigeria is either flared or re injected to enhance petroleum production [43]. In addition, vandalism of the gas network, for political reasons and in connection with theft from oil pipelines, has been blamed for much of the unreliability of the gas supply. [41] confirms that Nigeria does not make full use of its abundant gas resources for power generation because a lot of gas is flared or reinjected to enhance petroleum production. [43] also cites inadequate supply of gas to thermal stations due to high level of gas flaring and vandalism of pipelines as a major cause of underutilization of capacity.

We discuss the transmission network, installed generation capacity and the historical reliability of that capacity in more detail when we describe a model of the transmission generation system in Section 5.3.

2.4.5 Electricity Demand

As we have observed, the Nigerian power supply system fails to meet total electricity demand. In the context of a chronic power shortage we must distinguish between the electrical *load* on the power system and the *demand* for electricity. We adopt the following terminology to discuss difference between the actual load supplied and electricity demand. We refer to the level of demand that is met by the constrained electricity supply as the *suppressed demand* whereas the underlying level of electricity demand that customers connected to the grid would consume if supply was not constrained is the *unsuppressed demand*. The difference between unsuppressed demand and the suppressed demand is *unsatisfied demand*.

Load shedding causes a great deal of uncertainty about the demand for electricity because the unsuppressed demand is never measured. Several published forecasts seem quite optimistic. For instance, [45] present the results of forecasts of energy supply and demand for 2005 to 2030 based on International Atomic Energy Agency models. This forecast predicted peak demand of between 28 and 64 GW by 2015 and supply was forecast to be between 28 and 31 GW. TCN's estimated of peak load varies from 12,800 MW just after privatization to 14,630 MW in January 2016. A much lower figure was given by a load study commissioned by the Power Holding Company of Nigeria commissioned from the consultancy Tractabel, which reported in 2009. This study gave low, medium and high scenarios for the likely peak load values in 2010 of 4620, 4830 or 5100 MW respectively. In either case the potential load on the system is substantially in excess of its present capacity to deliver power.

Chapter 3

Demand Forecasting

3.1 Introduction

Demand modelling is of great importance for efficient operation, control and planning of power systems. However, in a chronic power shortage demand cannot be directly measured from observed load because this does not consider the unsatisfied demand due to load-shedding. In the context of a chronic power shortage we must distinguish between the electrical *load* on the power system and the *demand* for electricity. In order to balance the supply and load during a power shortage the Distribution System Operators (DSOs) constrain demand by shedding load (i.e. implementing rolling blackouts). We refer to the level of demand that is met by the constrained electricity supply as the *suppressed demand* whereas the underlying level of electricity demand that customers connected to the grid would consume if supply was not constrained is the *unsuppressed demand*. The difference between unsuppressed demand and the suppressed demand is *unsatisfied demand*. The objective of this chapter is develop methods for making short term forecasts of: 1) the electrical load on the power system conditional on particular load shedding decisions of the DSO (i.e. the suppressed demand); and 2) the latent demand level (i.e. the unsuppressed demand). In addition to forecasting we would also like to estimate the historic unsuppressed demand conditional on the observed suppressed demand. Our methods are agnostic to the cause of the power shortage which may be due to a shortage of generation capacity, transmission capacity or primary energy sources.

In many developing countries there is often a significant *unconnected demand* from consumers who are unable to get a grid connection or choose to meet their demand from off-grid sources because of reliability issues. Unconnected demand is outside the scope of this chapter. Furthermore, it is not the intention of this chapter to estimate what the electricity demand would be if load shedding were to be substantially reduced or eliminated. This would likely cause behavioural changes that are not accounted for by our models.

Standard methods for forecasting electricity demand cannot be straightforwardly applied to power systems in a state of chronic load shedding. During a period of load shedding only the suppressed demand, not the unsuppressed load can be measured directly. However, the suppressed load is determined both by the latent unsuppressed demand and by the load shedding decisions of the DSO. Therefore, it would not be adequate to merely apply standard electricity forecasting methods to the suppressed demand without in some way taking account of the load shedding policy.

This chapter proposes two forecasting approaches based on the Linear Guassian State Space Model (LGSSM). We test these methods using a new dataset obtained from a DSO operating in conditions of chronic power shortage. The average availability level of feeders in our data set is approximately 40%. We also know the number of customers supplied by each feeder according to the DSO's customer database. The customer numbers are broken down by tariff category into residential, commercial, industrial, street lighting and "special" customers.

We note that as in developed power systems the forecasting problem is likely to become more difficult with increasing penetration of distributed variable-output renewable energy sources. However, because of low renewable energy penetration it is unlikely that energy demand is yet materially affected by distributed renewables [3].

3.1.1 Electricity Demand Forecasting

The subject of short-term electricity demand forecasting has been widely studied. Popular procedures include ARIMA methods [13], the Kalman filter [32] and exponential smoothing [51]. A recent trend is the increased use of Neural networks which are appealing to model non-linear relationships between demand and weather variables. However, the advantages of Neural networks for short-term electricity forecasting unclear[13]. Another trend in the literature is towards probabilistic forecasts which aim to predict a probability distribution for future demand and provide a better quantification of uncertainty in addition to providing a point estimate [29]. For longer term forecasts it is important to capture the effect of more variables than the history of demand so approaches that model the relationship between electricity demand and exogenous variables (e.g. temperature) are popular. Some time series methods allow for exogenous regressor terms or explicitly model the evolution of the regressor variables and their correlation to electricity demand. However, [51] observes that multivariate modelling is usually considered impractical for real-time, on-line forecasting and univariate modelling is sufficiently accurate over short time scales because the effect of exogenous variables is captured in the demand series itself.

Medium to long term forecasts of electricity demand are often done using econometric methods that estimate the relationships between demand and exogenous variables like economic activity, population growth and climate and weather variables. This is not the direct aim of the models developed in this chapter. However, we note that a precondition for this is a reliable record of past electricity demand, but in a long-term power shortage electricity demand cannot be directly measured from the suppressed load. For example [34] and [2] investigate the relationship between electricity consumption and economic activity in Nigeria. Their efforts are hampered by the level of load shedding which means that the level of consumption is driven largely by supply rather than demand. Models which allow the historical unsuppressed demand to be estimated would therefore by of use.

A 2009 study, carried out on behalf of the Power Holding Company of Nigeria by the consultancy Tracrabel attempted to calculate the unsuppressed electricity demand using energy billings, injected energy and peak load as proxies. Two of their methods relied on extrapolating trends that are estimated using data before the Nigerian supply shortage became particularly acute. The inaccuracy of these methods should be expected to be substantial. Firstly because of the length of time elapsed since the data that was extrapolated. Secondly because this ignores the causal effect that long-term load shedding may have had in damping growth in electricity demand. The other methods also made dubious assumptions that there are no distribution system constraints and that the effect of load shedding only reduced load at peak hours. At best these methods can hope to give a coarse-grained estimate of the maximum load. They cannot give any useful guidance on the shape of the load curve throughout the day or the relationship between demand and factors such as temperature or economic variables.

3.1.2 Overview of Our Models

Load-shedding is carried out by disconnecting distribution feeders. This disconnection can take place at different levels in the network hierarchy. At a higher voltage level switching a single breaker might disconnect multiple feeders each of which can also be individually switched. The load measured at higher levels in the network hierarchy therefore depends on which lower level feeders are connected. When only a subset of the lower level feeders are connected the measured load does not include demand from the disconnected feeders. At a sufficiently low level in the network hierarchy there are breakers which control only a single feeder. In this chapter I will use the term *distribution feeder* to refer only to a sub-branch of the distribution network below which there is no further load-shedding.

In our Agreggated Structural Models (ASM) approach we assume that when a distribution feeder is connected, the recorded load gives a measure of the unsuppressed demand on that feeder. The fundamental problem is therefore one of modelling time series with a high degree of missing data. The state space approach to time series modelling using the Kalman filter is well suited to time series with a lot of missing data. In the state space approach to time series modelling we assume the time series is composed of various unobserved components and posit and explicit stochastic model for each of them. If the model is linear and Gaussian, then the Kalman filter can be applied to the state space representation to efficiently calculate the maximum likelihood estimates of the unobserved components conditional on the observed data and the assumed stochastic model. The parameters of the stochastic model can then be optimized to maximize the likelihood.

An appropriate state space model to represent electricity demand is the so called "structural model" in which the observed series is assumed to be composed of unobserved stochastic processes representing periodic "seasonal" effects and variation in the local "level" of the series and a random observation error. In the short-term electricity demand is subject to transient local variations due to, for example, metrological and social phenomena as well as periodic variations over daily, weekly and longer timescales. These phenomena are accounted for by the structural model. We can therefore model the unsuppressed electricity demand on each feeder by a structural model and sum the estimated loads on the connected feeders to estimate the suppressed load. Hence, we refer to this as the Aggregated Structural Models approach. The underlying assumption is that the observation errors and stochastic evolution of each feeder is independent of the others. Over the short-term (i.e. several hours) this may be a reasonable approximation. We discuss how we can relax this assumption in Section 3.3.4.

Our approach directly models non stationary features that are inherent to electricity demand data and performs interpolation of the missing data. The alternative approach of first removing non stationary features by differencing and then modelling the residuals as a stationary series is problematic in our case. In order to difference the time series, we would have to first interpolate the data gaps. The quantity of missing data means that the resulting model would give a good fit to the interpolated data at the expense of the real data.

Our Dynamic Structural Regression (DSR) approach is motivated by the observation that rather than modelling each individual feeder we might prefer to model the unsuppressed demand of the whole DSO with a structural model. Of course, we have no observations of unsuppressed demand to fit this model. However, if we assume that customers are homogeneous then the suppressed demand is given by multiplying the unsuppressed demand by the proportion of total customers connected. The level of unsuppressed demand then plays the role of a time varying regression coefficient which can be estimated using the Kalman Filter. We propose a model in which the level of unsuppressed demand is assumed to follow the dynamics of the Structural Model. We can form a State-Space model where an unobserved state variable, the unsuppressed demand, is multiplied by an exogenous regression variable, the proportion of customers connected, to give the mean of the observation distribution. We can relax the assumption of customer homogeneity by breaking down the customer numbers into different categories (e.g. residential and industrial) and forming a different model for each type of customer. This allows us to better model the effects of connecting feeders with different mixtures of customer types. In summary this model is a dynamic regression model in which the regression coefficients are assumed to vary over time according to a Structural Model, hence the name Dynamic Structural Regression.

Time varying parameter models in which unobserved regression coefficients evolve according to a random walk or stationary process have been studied in the econometrics literature (for example in [12] and [24]). However, to the best of our knowledge the DSR model has not been previously studied. By using a structural model for the coefficients our model is better able to handle periodic variation in the coefficients. This model applies to the general case where a time series of observations are given by the product of a known level of "measurement effort" and an underlying process with stochastic dynamics that can be well approximated by a structural model. An alternative way of handling periodicity in dynamic regression models is with dynamic factor models as in[15]. This involves transforming the time series into a lower frequency multivariate time series. We do not pursue this because our aim is to compare the customer regression based approach exemplified by the DSR model to the feeder bases approach exemplified by the ASM approach. An interesting topic for future work would be to compare how different ways of incorporating periodicity into dynamic regression models compare for signal extraction and forecasting purposes.

In the next section we provide a detailed description of our methods. We first review the general LGSSM before going on to describe how our ASM and DSR models are represented in state space form. We then discus two modifications to the models to account for autocorrelated errors and possible correlations between individual feeders respectively. In the subsequent section we carry out an empirical comparison of our methods using a real dataset of feeder level electricity demand and customer numbers obtained from a DSO operating under conditions of continual load shedding. We first consider the accuracy of point forecasts and then theoretical forecast error distributions of our models.

3.2 Theoretical Background

3.2.1 The Linear Gaussian State Space Model

Consider a discrete time stochastic process Y with individual elements y_i. The basic time series problem is to infer the distribution (or some other statistical property such as the expectation) of $y_t \in Y$ given a set of noisy measurements $\{y_i : i \in \{1,2,...\tau\}\}$. If $\tau < t$ we call this problem forecasting, if $\tau = t$ we call the problem filtering and if $\tau > t$ we call the problem smoothing.

For a time series model in state space form the observation y_t is assumed to be a linear combination of the elements of a state vector α_t plus a normally distributed observation error ε_t.

The relationship between the state at time t and $t+1$ is given by a deterministic linear function plus a vector of normally distributed state disturbances η_t.

The model can be summarized in matrix form:

$$y_t = Z_t \alpha_t + \varepsilon_t, \tag{3.1}$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \tag{3.2}$$

with $\varepsilon_t \sim N(0, H_t)$ and $\eta \sim N(0, Q_t)$

Equation 3.1 is the *observation equation* and Equation 3.2 is the *state transition equation.*

In order specify a LGSSM we must define the design matrix Z_t, state transition matrix T_t, selection matrix R_t, observation error covariance matrix H_t and state disturbance covariance matrix Q_t. In the general case these matrices can have time varying components, hence they are indexed by t. The model is called *time-invariant* if the state space models do not vary with time.

Why We Do Not Use the "Innovations" State Space Model

Two approaches to formulating state space models are distinguished by how the state is assumed to evolve over time. In the conventional state space approach, the state is assumed to evolve according to a stochastic process which is independent of the observation errors. By contrast in the "innovations" approach the state is assumed to evolve in a deterministic fashion conditional on the value of the forecast error, so the next state can be updated exactly. As a result, innovations State Space models are computationally easier to fit because they do not require the Kalman filter. The innovations approach is extensively developed in [31]. These models are shown to provide a theoretical justification for commonly used "exponential smoothing" forecasting procedures. Exponential smoothing methods have been applied to electricity demand in [51] and [52]. We choose to focus on the conventional state space approach using the Kalman filter in this chapter because of the sound theoretical treatment of missing data by the conventional state space approach and the ready availability of software for forming custom state space models. Henceforth we use the term State Space model to refer to the conventional State Space approach using multiple independent sources of error and the Kalman filter.

3.2.2 The Kalman Filter and Smoother

Given a LGSSM and a known initial state (or known probability distribution of the initial state) and known variances for the state disturbance terms and observation error we can compute the maximum likelihood estimates for the state at time $t+1$ conditional on all the prior data, $a_{t+1} = E[\alpha_{t+1}|y_1, ...y_t]$ using the Kalman filter.

Let the forecast error at time t be given by $v_t = y_t - Z_t a_t$ and suppose that the state estimation error covariance matrix P_t is known. The Kalman Filtering recursion is given by the equations:

$$F_t = Z_t P_t Z_t' + H_t, \tag{3.3}$$
$$L_t = T_t - K_t Z_t, \tag{3.4}$$
$$K_t = T_t P_t Z_t' F_t^{-1}, \tag{3.5}$$
$$a_{t+1} = T_t a_t + K_t v_t, \tag{3.6}$$
$$P_{t+1} = T_t P_t L_t' + R_t Q_t R_t' \tag{3.7}$$

Equation 3.3 is the forecast error covariance matrix, 3.4 is used as part of 3.5 to calculate the Kalman gain matrix. These are used to update the estimated state in 3.6. 3.7 updates the state estimation covariance matrix.

Note that a_t, P_t and F_t are calculated based on all observations available before time t. We will use the notation $X_{t|s}$ to mean the estimate of X at time t given information up to and including time s. We adopt the convention that $X_t = X_{t|t-1}$.

Initializing the Kalman Filter

The Kalman filter defined the estimated state a_t and state estimation error covariance P_t recursively. Therefore, some initial values must be specified.

The Kalman filter has its origins in engineering applications where it is often practical to initialize with a known state, or a known mean and variance. In most time series applications this is not practical, except for stationary time series where the unconditional mean and variance can be derived. Where this is not possible a common approach is to use “diffuse” initialization. This involves making the assumption that the initial states have infinite variance. This requires some modifications to be made to the Kalman Filter algorithm, but it can be approximated by using a very large variance relative to the variance of the error terms. This is the approach implemented in the python package statsmodels [46]. Because of the approximate diffuse initialization, we discard the first k data points after initialization when assessing the goodness of fit or forecasting performance of the models. k is typically set to the dimension of the state vector. This is a practical approach to getting rid of the effect of initialization for our data because of the length of the available data. For shorter series it would be more important to use exact diffuse initialization.

Parameter Optimization

An application of the Kalman filter implies particular values for the observation and disturbance terms. Since these are serially independent (by the assumptions of the model) and have known variance we can easily evaluate the likelihood of the data given particular parameter values for the variance of the state disturbances and observation errors. It is then possible to optimize the parameter values to maximize the likelihood.

Forecasting With the Kalman Filter

Suppose we have a time series which we have observed up to time s and we want to forecast up to some forecast horizon h. That is, we want forecasts for all times $t : s < t < t+h$. Let's assume we have used the Kalman filter to compute the estimated state a_s and the state estimation error covariance matrix P_s.

We can forecast the time series by computing the expected value of the state, α_t, and multiplying by Z_t. To do this we project the estimated state a_t forwards recursively by multiplying the estimated state by the transition matrix T_t to get the expected value of the state at time $t+1$

Forecast uncertainty

The error of our forecast will be given by the error in our estimation of the state and the observation noise. The contribution to the forecast error from the state estimation error is given by $Z_t P_t Z_t'$. This is independent of the observation noise which has known variance denoted H_t. The Forecast error covariance matrix is therefore given by $F_t = Z_t P_t Z_t' + H_t$.

3.3 Methods

3.3.1 Aggregated Structural Models (ASM)

Here we define the state space representation of the structural model that we use to model each feeder.

We define the observed demand on feeder i, $d_{i,t}$ to be the sum of two state variables, $\mu_{i,t}$ and $\gamma_{i,t}$ and a normally distributed error $\varepsilon_{i,t} \sim N(0, \sigma_\varepsilon^2)$.

$\mu_{i,t}$ is defined so as to model local trends and occasional shocks to the level of the series using a simple random walk model:

$\mu_{i,t+1} = \mu_{i,t} + \xi_{i,t}$

where $\xi_{i,t} \sim N(0, \sigma_\xi^2)$

$\gamma_{i,t}$ models "seasonal" effects that repeat with a period of constant length. Let s be the number of time periods in a single seasonal cycle, e.g. $s = 24$ for hourly data and daily seasonality. We can use either a *time domain* or *frequency domain* model for the seasonal component.

The time domain approach models the seasonal effects by a unique additive effect for each period within the seasonal cycle which are required to sum to 0. We can calculate the seasonal term at time t from the last $s-1$ seasonal terms: $\gamma_{i,t+1} = -\sum_{j=1}^{s-1} \gamma_{i,t+1-j}$. Allowing the seasonal terms to evolve over time is a simple matter of adding an error variable $\omega_{i,t} \sim N(0, \sigma_\omega^2)$ to the right hand side of this equation.

The frequency domain approach involves expressing the seasonal component as a sum of trigonometric terms at frequencies $\lambda_j = \frac{2\pi j}{s}$ for $j = 1, ..., [s/2]$:

$$\gamma_{i,t} = \sum_{j=1}^{[s/2]} \gamma_{i,j,t}$$

where:

$$\gamma_{i,j,t+1} = \gamma_{i,j,t}\cos(\lambda_j) + \gamma^*_{i,j,t}\sin(\lambda_j) + \omega_{i,t}$$

$$\gamma^*_{i,j,t+1} = -\gamma_{i,j,t}\sin(\lambda_j) + \gamma^*_{i,j,t}\cos(\lambda_j) + \omega^*_{i,t}$$

$$\omega_{i,t} \sim N(0, \sigma^2_{\omega_i}) \text{ and } \omega^*_{i,t} \sim N(0, \sigma^2_{\omega^*_i})$$

If the full range of frequencies is included then the frequency domain model is identical to the time domain model, however it is possible to drop some of the frequencies to obtain a more parsimonious model for complex seasonal patterns. For example, suppose we are working with hourly data that displays a daily pattern and weekly pattern. It may be possible to model the weekly pattern with just the frequencies for $j = 1, 2, 3$ and the daily pattern with just the frequencies $j = 7, 8$. It is common to set the variances equal for the disturbance terms at all frequencies j, however, in principle they can be allowed to vary. In the example just discussed it might be desirable to have a different variance for the frequencies modelling the weekly seasonality to those modelling daily seasonality. For these reasons we prefer to use the frequency domain approach.

Putting this all together we get the observation equation:

$$d_{i,t} = \mu_{i,t} + \gamma_{i,t} + \varepsilon_{i,t}$$

And the following transition equations:

$$\mu_{i,t+1} = \mu_{i,t} + \xi_{i,t}$$

$$\gamma_{i,t} = \sum_{j=1}^{[s/2]} \gamma_{i,j,t}$$

$$\gamma_{i,j,t+1} = \gamma_{i,j,t}\cos(\lambda_j) + \gamma^*_{i,j,t}\sin(\lambda_j) + \omega_{i,j,t}$$

$$\gamma^*_{i,j,t+1} = -\gamma_{i,j,t}\sin(\lambda_j) + \gamma^*_{i,j,t}\cos(\lambda_j) + \omega^*_{i,j,t}$$

and $\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$, $\xi_t \sim N(0,\sigma_\xi^2)$, $\omega_t \sim N(0,\sigma_\omega^2)$ and $\omega^*_{i,j,t} \sim N(0,\sigma^2_{\omega^*_j})$

The matrix form of the basic structural model is well known and is documented, for example, in [17]. We now define the matrices required to define the dynamic regression model in state space form.

3.3.2 Dynamic Structural Regression (DSR)

Let d_t denote the total demand from all feeders connected at time t and let $n_{x,t}$ denote the proportion of customers of type x that are connected at time t. Let us assume that there are four categories of customers and that the indices R,C,D and S stand for residential, commercial, industrial and special, and street lighting customers respectively.

A multiple linear regression model for d_t is given by:

$$d_t = \beta_R n_{R,t} + \beta_C n_{C,t} + \beta_D n_{D,t} + \beta_S n_{S,t} + \varepsilon_t$$

$$\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$$

where β_x is the regression coefficient for customer type $x \in \{R,C,D,S\}$.

The regression coefficients can be interpreted as estimates of total unsuppressed demand for all customers of each customer category. This multiple linear regression model assumes that the unsuppressed demand remains constant over time. This is an implausible assumption because of periodic variation at various time scales and because of transient variation in the level of demand due to, for example, changing weather patterns.

To account for this, we propose a dynamic regression model in which the regression coefficients have a stochastically evolving mean and periodic component:

$$d_t = \beta_{R,t} n_{R,t} + \beta_{C,t} n_{C,t} + \beta_{D,t} n_{D,t} + \beta_{S,t} n_{S,t} + \varepsilon_t$$

$$\beta_{x,t} = \mu_{x,t} + \gamma_{x,t},\ \forall x \in \{R,C,D,S\}$$

$$\mu_{x,t} = \mu_{x,t-1} + \eta_{x,t}$$

$$\gamma_{x,t} = \sum_{j=1}^{[s/2]} \gamma_{j,x,t}$$

where:

$$\gamma_{j,x,t+1} = \gamma_{j,x,t}\cos(\lambda_j) + \gamma^*_{j,x,t}\sin(\lambda_j) + \omega_{j,x,t}$$

$$\gamma^*_{j,x,t+1} = -\gamma_{j,x,t}\sin(\lambda_j) + \gamma^*_{j,x,t}\cos(\lambda_j) + \omega_{j,x,t}$$

$$\lambda_j = \frac{2\pi j}{s} \text{ for } j = 1,\ldots,[s/2]$$

and $\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$, $\eta_{x,t} \sim N(0,\sigma_\eta^2)$ and $\omega_{j,x,t} \sim N(0,\sigma_\omega^2)$

Matrix Form of Dynamic Structural Regression Model

We wish to define a design matrix Z_t and state vector α_t such that $Z_t\alpha_t = \beta_{R,t}n_{R,t} + \beta_{C,t}n_{C,t} + \beta_{D,t}n_{D,t} + \beta_{S,t}n_{S,t}$ where $\beta_{x,t}$ is defined as in the section above.

The state vector can be divided into two sets of state variables:

$\alpha_t = (\boldsymbol{\mu}_t, \gamma_t)^T$

The first part of the state vector gives the current level of the random walk for each regression coefficient:

$\boldsymbol{\mu}_t = (\mu_{R,t}, \mu_{C,t}, \mu_{D,t}, \mu_{S,t})$

The second part contains the seasonal components of each regression coefficient:

$\gamma_t = (\gamma_t^R, \gamma_t^C, \gamma_t^D, \gamma_t^S)$

The regression coefficient for customer category $x \in \{R, C, D, S\}$ has 2 state variables for each harmonic frequency, h, included in the model: $\gamma_{h,t}^x$ and $\gamma_{h,t}^{x*}$.

The seasonal components of each regression coefficient are given by concatenating these state variables. Assuming the the full range of harmonics for 24-hour seasonality is included this gives 24 state variables for each customer category:

$\gamma_t^x = (\gamma_{1,t}^x, \gamma_{1,t}^{x*}, ..., \gamma_{12,t}^x, \gamma_{12,t}^{x*},)$

Let $n_{x,t}$ denote the proportion of customers of category $x \in \{R, C, D, S\}$ connected at time t and $\boldsymbol{N}_{x,t}$ denote the vector containing 12 repetitions of $(n_{x,t}, 0)$ (i.e. one for each harmonic).

Define design matrix Z_t as the vector $(n_{R,t}, n_{C,t}, n_{D,t}, n_{S,t}, \boldsymbol{N}_{R,t}, \boldsymbol{N}_{C,t}, \boldsymbol{N}_{D,t}, \boldsymbol{N}_{S,t})$

Then we have:

$$Z_t\alpha_t = \sum_{x\in\{R,C,D,S\}} (\mu_{x,t} + \sum_{h\in 1,...,12} \gamma_{h,t}^x) n_{x,t} = \beta_{R,t}n_{R,t} + \beta_{C,t}n_{C,t} + \beta_{D,t}n_{D,t} + \beta_{S,t}n_{S,t}$$

where $\beta_{x,t} = \mu_{x,t} + \gamma_{x,t}$, $\forall x \in \{R, C, D, S\}$, as required.

We now wish to define a transition matrix T, vector of disturbance terms η_t and selection matrix R in such a way that $T\alpha_t + R\eta_t = \alpha_{t+1}$ correctly updates $\mu_{x,t}$ and $\gamma_{x,t}$ to $\mu_{x,t+1}$ and $\gamma_{x,t+1}$ according to the relations defined in the previous section.

Let us define $C = \text{diag}[C_1, ..., C_{12}]$ with $C_j = \begin{bmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{bmatrix}$ and $\lambda_j = \frac{2\pi j}{s}$ for $j = 1, ..., 12$.

The vector of disturbance terms is given by $\eta_t = (\eta_{R,t}, \eta_{C,t}, \eta_{D,t}, \eta_{A,t}, \omega_{j,R,t}, \omega_{j,C,t}, \omega_{j,D,t}, \omega_{j,S,t})^T$

Let $\boldsymbol{e}_n$ denote the n element column vector of ones.

Then if we define $R = \text{diag}[I_4, \boldsymbol{e}_{24}, \boldsymbol{e}_{24}, \boldsymbol{e}_{24}, \boldsymbol{e}_{24}]$ and $T = \text{diag}[I_4, C, C, C, C]$ we will have $T\alpha_t + R\eta_t = \alpha_{t+1}$ as required.

3.3.3 Accounting for Autocorrelated Errors

The LGSSM assumes that the observation errors and state disturbance terms are serially independent. If this assumption is not satisfied, then a possible fix is to supplement both the models above by adding states that evolve according to stationary autoregressive processes.

Let ψ_t denote a process which evolves according to a first order autoregressive process with autoregressive parameter Θ and error variance σ_ψ.

We replace the observation equation of the structural model for each feeder with:

$d_{i,t} = \mu_{i,t} + \gamma_{i,t} + \psi_{i,t} + \varepsilon_{i,t}$

In the DSR case we replace the equation defining the regression coefficient for customer type x with:

$\beta_{x,t} = \mu_{x,t} + \gamma_{x,t} + \psi_{x,t}, \forall x \in \{R, C, D, S\}$

There are various ways that an autoregressive process can be represented in state space form; we adopt the method of [26].

3.3.4 Accounting for Correlation of Errors Between Feeders

The ASM approach assumes that the errors of the structural models for each feeder are independent. In practice this is unlikely to be true. It is well known that electricity consumption is correlated with weather patterns and social events which are likely to be correlated between feeders.

One approach for handling this is to replace the observation equation for each of the n feeder models with the single equation $\mathbf{y}_t = Z\alpha_t + \varepsilon_t$ where $\mathbf{y}_t$ is an n dimensional vector of demand observations and α_t is the state vector formed by concatenating the states of all n independent feeder models. ε_t is an n dimensional vector of observation errors with multivariate normal distribution: $\varepsilon_t \sim N(0, H)$ where H is an n by n covariance matrix. If we have p states, then the pn dimensional state disturbance vector is multivariate normally distributed with a pn by pn covariance matrix. The deterministic part of the evolution of the state for each feeder is therefore independent of every other feeder, but the observation errors and state disturbance terms are correlated. [17] refered to this model as "seemingly unrelated time series equations".

However, this requires the estimation of an n by n covariance matrix H and pn by pn covariance matrix Q which is impractical given the number of feeders we must model. Therefore, we propose a more computationally tractable, two-stage approach by combining our ASM and DSR methods.

We first use the DSR model to arrive at an estimate of unsuppressed demand for all the feeders. We then supplement all the structural models in the ASM approach with a linear regression term with the estimated unsuppressed demand from the DSR model as the exogenous covariate.

The effect of this is that each of the structural models in the ASM approach models the difference between the demand on that feeder and what would be expected given the estimated overall demand. Thus, the assumption of independence between the feeder observation errors is replaced with an assumption of independence conditional on the level of overall demand estimated by the DSR model.

3.3.5 Theoretical Forecast Error Variance

According to the ASM model the total demand is the sum of the demand on the connected feeders. Let $x_{i,t}$ be a binary variable taking the value 1 at time t if feeder i is connected, 0 otherwise. Total demand is therefore given by:

$$d_t = \sum_{i,t} x_{i,t} d_{i,t}$$

Where $d_{i,t}$ is the demand on feeder i modelled by the state space observation equation:

$$d_{i,t} = Z_{i,t}\alpha_{i,t} + \varepsilon_{i,t}$$

The state $\alpha_{i,t}$ is not known so at time $s < t$ we estimate it with using the Kalman filter by $a_{i,t|s}$. Our forecast demand is therefore given by $\sum_{i,t} x_{i,t} Z_{i,t} a_{i,t|s}$. Under the assumptions of the model the true demand is distributed around this forecast according to the distribution of the sum of the forecast errors.

According to the DSR model the proportion of connected customers of type i connected at time t is $n_{i,t}$. The total demand from customers of type i is given by the coefficient $\beta_{i,t}$. The total demand is $d_t = \sum_i n_{i,t}\beta_{i,t} + \varepsilon_t$

This can be represented in state space matrix form as shown in Section 3.3.2 by defining the design matrix as a function of the vector of connected customers of each type, $Z_{i,t} = Z(n_t)$. The coefficients for each customer type are contained in the state vector α_t, estimated at time $s < t$ by $a_{i,t|s}$. The forecast demand from the DSR method is therefore given by $Z_{i,t}\alpha_{i,t|s}$ and is normally distributed around this with variance given by $F_t = Z(n_t)P_t Z(n_t)' + H_t$. All other things being equal he forecast variance increases quadratically with $n_{i,t}$.

For both ASM and DSR the forecast variance is higher when more feeders are connected. In practice this tends to mean that the variance is highest at night because more feeders can be connected at off-peak hours.

Multiplicative Models

It is sometimes more plausible to think that a time series has greater variance at times when the series has a high value. In the context of electricity demand we might think that the peak demand level has more uncertainty than off-peak demand. One way to handle this is to specify a model structure so that the errors are proportional to the level of the time series

Consider the following multiplicative model where y_t is the observed value, α_t and β_t are stochastically evolving states and ε_t is an independent error term.

$$y_t = (\alpha_t + \beta_t + \gamma_t) * \varepsilon_t$$

This model is non-linear so we cannot put it in the form of a LGSSM and apply the Kalman filter to estimate the states. However, if the states and observation errors are combined multiplicatively then we can lake the natural logarithm of both sides to obtain an additive decomposition as follows:

$$y_t = \alpha_t * \beta_t * \gamma_t * \varepsilon_t$$

$$ln(y_t) = ln(\alpha_t * \beta_t * \gamma_t * \varepsilon_t)$$

$$ln(y_t) = ln(\alpha_t) + ln(\beta_t) + ln(\gamma_t) + \ln(\varepsilon_t)$$

The additively decomposed form can be modelled by a LGSSM which implies that ε_t has a log-normal distribution.

Recall the structural model defined in Section 3.3.1. By taking logarithms we obtain a model in which the mean, seasonal component and error combine multiplicatively and for which the values of the time series are restricted to the positive real numbers.

This modelling approach is much less appealing as a modification of the DSR model developed in Section 3.3.2.

Suppose we model $ln(d_t)$ by the multiple linear regression model:

$$ln(d_t) = \beta_R n_{R,t} + \beta_C n_{C,t} + \beta_D n_{D,t} + \beta_S n_{S,t} + \varepsilon_t$$

It follows that

$$d_t = e(\beta_R n_{R,t}) + e(\beta_C n_{C,t}) + e(\beta_D n_{D,t}) + e(\beta_S n_{S,t}) + e(\varepsilon_t)$$

which implies a non-linear relationship between the proportion of each customer type connected and demand which is not a plausible model.

An alternative approach is to exploit the fact that the LGSSM allows the variance of the observation error and state disturbances to vary over time. This variation over time is exogenous to the model. Therefore, it is possible in principle to define variance as a function of time in such a way that the variance exhibits a periodic pattern and is higher during peak hours. We can fit the parameters of such a function at the same time as fitting the parameters of the LGSSM.

3.4 Empirical Comparison of Methods

In this section we evaluate and compare our models by applying them to a dataset obtained from a Nigerian DisCo. The dataset consists of hourly records of the load on a set of distribution feeders that have been manually recorded by distribution substation operators. At the beginning of every hour a numerical value is recorded if the feeder is energized. Otherwise the status of the feeder is recorded with a character string representing events such as load shedding (L/S) or system outages (S/O). In order to fit the DSR models (and the ASM models which use the DSR output as exogenous covariates) we need to supplement this data with the proportion of customers of each type that are connected in every hour. In the Nigerian case customers are divided by type and size into 14 categories as shown in Table 3.1. We have access to a dataset giving the number of active customers supplied by each feeder on a monthly basis split by tariff categories. We can interpolate this data to create an hourly dataset. We then take the sum over all the connected feeders to calculate the number of connected customers and then normalize this by dividing by the total number of customers.

Jos Electricity Distribution Company (JEDC) operates a large franchise area in the North East of Nigeria. By comparison with larger distribution companies covering Nigeria's major metropolitan areas its distribution network is less dense and connects fewer customers. The dataset contains data from 76 distribution feeders and is available from January 2017 to May 2018. However, for reasons of practicality we restrict our attention to the period starting in the first hour of 2nd January up to and ending after the last hour of the 1st April 2017. This is sufficient data to fit the models without over fitting (As we demonstrate in the out of sample analysis in Section 3.4.3) and means we can fit the models in a practical amount of time. A further benefit of limiting the scope is that this period does not contain any public holidays or other special days, so it avoids the need to account for these within the model (Although this can be easily achieved in the state space framework with dummy intervention variables).

3.4.1 Model Building

We have described two general families of model, the ASM and DSR model. The ASM model relies upon fitting a Structural Model to each feeder. The DSR approach models demand at a more aggregated level by modelling the relationship between the number of connected customers and the resultant demand as a dynamic regression model. Within each model family there are some further modelling decisions to be made.

The number of potential models within each model type is quite large. It is not practical to exhaustively fit and compare all the potential model configurations. We study the results of fitting a basic version of the ASM and DSR models using a model of the daily effects with 5 seasonal harmonic frequencies. In addition we try supplementing each model with and autoregressive order 1 model as described in Section 3.3.3. Finally, we include the unsuppressed demand as

Customer Classification	Number	Description	Rermarks
Residential			
R1	431	Life-Line (50 kWh)	A consumer who uses his premises exclusively as a residence- house, flat or multi-storeyed house
R2	162333	Single and 3-phase	
R3	108	LV Maximum Demand	
R4	8	HV Maximum Demand (11/33KV)	
Commercial			
C1	20158	Single and 3-phase	A consumer who uses his premises for any purpose other than exclusively as a residence or as a factory for manufacturing goods
C2	585	LV Maximum Demand	
C3	8	HV Maximum Demand (11/33KV)	
Industrial			
D1	1429	Single and 3-phase	A consumer who uses his premises for manufacturing goods including welding and ironmongery
D2	1	LV Maximum Demand	
D3	4	HV Maximum Demand (11/33KV)	
Special			
A1	945	Single and 3-phase	Customers such as agriculture and agro-allied industries, water boards, religious houses, government and teaching hospitals, government research institutes and educational establishments
A2	88	LV Maximum Demand	
A3	67	HV Maximum Demand (11/33KV)	
Street Lighting			
S1	15	Single and 3-phase	

Table 3.1: Nigerian Customer Tariff Categories and Customer Numbers in Jos EEDC at start of year 2018

estimated by the DSR model with and autoregressive error as an exogenous regression variable in the ASM approach. This gives 6 models, each of which we associate with a short acronym:

ASM - Aggregated Structural Models

ASM_ar - Aggregated Structural Models with AR(1) state

ASM_corr - Aggregated Structural Models with exogenous covariate from DSM_ar model

ASM_corr_ar - Aggregated Structural Models with exogenous covariate from DSM_ar model and AR(1) state for each regression coefficient

DSR - Dynamic Structural Regression

DSR_ar - Dynamic Structural Regression with AR(1) state for each regression coefficient

Representation of Daily and Weekly Seasonal Effects

For both model types we can choose whether to include a model of daily seasonal effects and if so, which harmonic frequencies of 12 to include in the model. Similarly, there are potentially 84 harmonic frequencies that could be included in a model of weekly seasonal effects. Arguably only the largest 6 of these that have a period longer than one day should be considered since effects with a shorter period will be captured in the model of daily seasonality.

The choice of 5 seasonal harmonics for the daily seasonal model and no representation of weekly seasonality was based on some preliminary experiments where we used the Akaike Information Criterion (AIC) to select the best fitting model. The AIC is a commonly used goodness of fit statistic used to compare models with different numbers of parameters. For a state space model where ω is the dimension of the parameter vector $\hat{\psi}$ and q is the number of diffuse initialized states the AIC is defined thus: $\text{AIC} = \frac{-2\log L(y|\hat{\psi})+2(q+\omega)}{n}$ [17]. This has the effect of penalizing models with more parameters and diffuse states.

Selection of Covariates for the DSR Model

For the DSR model we have some options with respect to how we group customers into different customer classes. It is not practical to infer dynamic regression coefficients for all 14 customer categories, therefore we need to aggregate the customer categories together somehow. The number of possible aggregations is very large. We will simplify this modelling choice by just considering the following grouping of customers:

Group all residential customers together and then group the remaining customers by size. Residential: R1, R2, R3, R4, S1; Small: C1, D1, A1; Medium: C2, D2, A2; Large: C3, D3, A3.

This selection of customers was based on some preliminary experiments where we grouped customers by type rather than size as follows:

Group all customers by type. Residential: R1, R2, R3, R4, S1; Commercial: C1, C2, C3; Industrial: D1, D2, D3; Special: A1, A2, A3.

We observed that the coefficients for the commercial, industrial and special categories were often quite erratic suggesting their is not the required level of customer homogeneity in these categories.

3.4.2 Model Fit

We first fit our models to 1080 hours of data and evaluate how well the modelling assumptions of the LGSSM are satisfied. The assumptions of the LGSSM are that the state and observation disturbances are normally distributed, serially independent and have the same variance at all time periods (homoskedasticity). Durbin and Koopman [17] suggest visually assessing the following plots of the standardized forecast residuals to test these assumptions: a plot of the standardized residuals over time; a histogram of the standardised residuals compared to its kernel density estimate and the standard normal Probability Density Function (PDF); A normal Quantile-Quantile (QQ) plot of the standardised residuals; and the autocorrelogram of the standardized residuals.

First, we consider the basic ASM model. A representative example of these diagnostic plots for 4 arbitrarily selected feeder is given in Figure 3.1. Diagnostic plots for all feeders are available in Appendix 9.

The time plots of the standardized residuals for all feeders exhibit large "spikes" indicative of outliers or structural breaks. In addition, for some of the feeders the size of the errors seems not to be similarly distributed over time, meaning that the homoskedasticity assumptions are not well satisfied. For example, the first feeder seems to have higher variance in the middle of the series which suggests a structural break occurred in the underlying data generating process.

The normal QQ plot should match the line $y = x$ if the standardized residuals are normally distributed. For most of the feeders this plot is strongly indicative that the forecast errors are not in fact normal. This could either mean that the model structure is flawed or that the disturbances are not normal. The general shape is indicative that the standardized forecast residuals are heavier tailed than the normal distribution. However, feeder 4 in our example is an exception; the QQ plot closely matches the line $y = x$ which is indicative that the standardized residuals are close to normally distributed.

However, looking at the kernel density estimate what is most apparent is that more probability mass is concentrated very close to the mean of the distribution. The autocorrelogram is suggestive of some residual autocorrelation which motivated us to consider adding an AR(1) error.

In Figure 3.2 we plot the observed values and one-step-ahead predictions for a representative examples of feeders. Similar plots for all feeders are available in the appendix. These plots

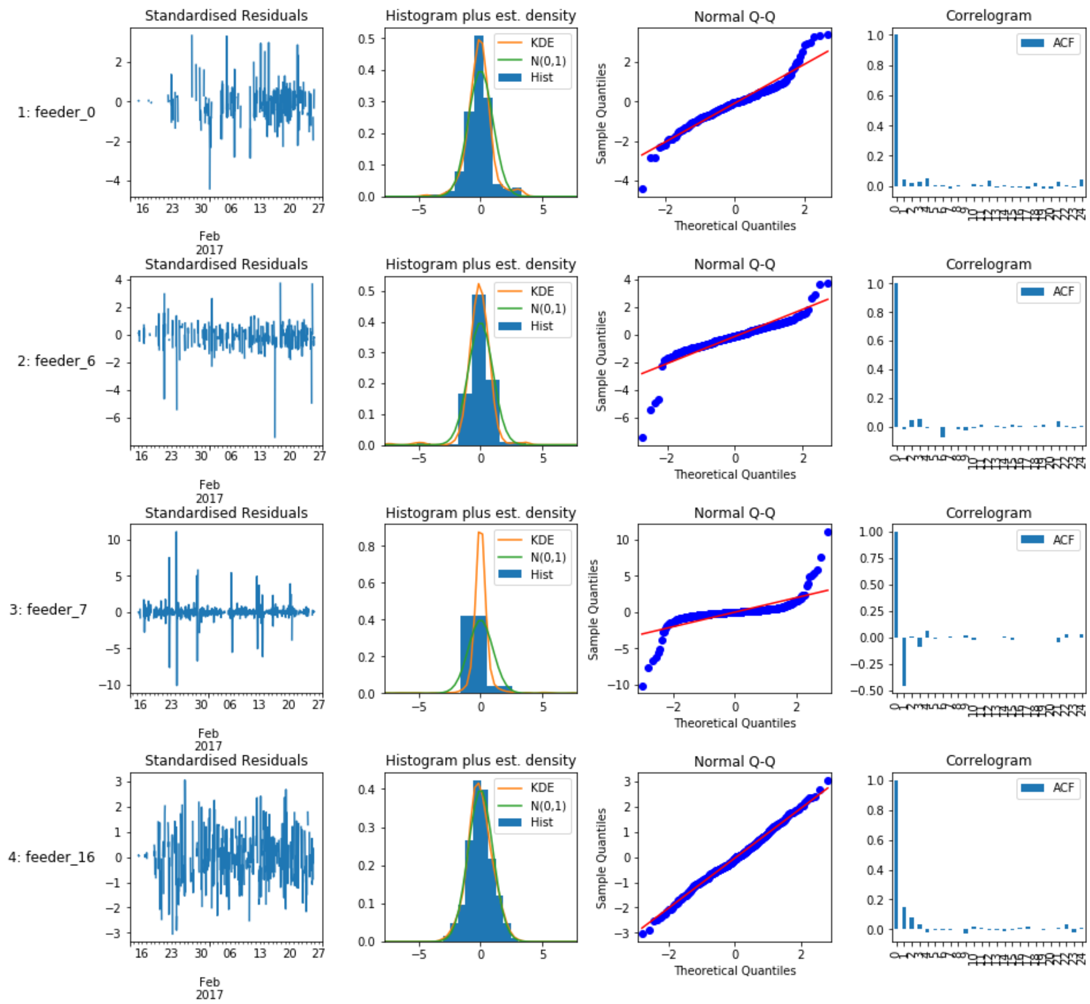


Figure 3.1: Diagnostic plots of structural models for a selection of feeders

illustrate the presence of outliers and structural breaks. We speculate that the structural breaks are caused by faults in the network that result in the disconnection of a fraction of the customers served by a given feeder. In principle a DisCo could gather this information and use dummy intervention variables to indicate when a fault occurred which may improve the model performance.

Diagnostic plots for the DSR model are given in Figure 3.3. By comparison with the structural models the standardized forecast residuals for the DSR model are closer to a normal distribution although the normal QQ plot still gives reason to believe that the residuals have a heavier tailed distribution than the normal distribution. The plot of the standardized residuals shows 4 particularly large values relative to the other observations.

3.4.3 Forecast Performance

For practical purposes it is useful to know the forecasting performance of our models for different forecast lead times. The longer the lead time for which accurate forecasts can be provided then the further ahead that distribution system operators can use them for planning. In the context of power systems in many developing countries DSOs lack the funds to install automated Supervisory Control and Data Aquisition (SCADA) systems and unreliable communications infrastructure and procedures mean that it may be difficult to obtain demand information in real time. As a result, forecasts even for the hour immediately ahead may in practice be multi-step forecasts because they have to be made without up-to-date information. We therefore compute forecast performance statistics for a forecast lead time up to 24 hours.

To compare models with different numbers of parameters it is common to compute statistics such as the AIC which penalize models with more parameters. In our case this is problematic because the ASM approach is fitted to a dataset containing more information by modelling each individual feeder time series. Therefore, rather than computing the AIC our approach is to perform an out of sample analysis.

Common measures of forecasting performance are the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE). We study both the MAPE and MAE statistic to assess the forecast performance of our models. A major advantage of the MASE over the MAPE is that it allows forecast comparisons between different time series, but that is not necessary for our purposes. The MAPE is undefined when the true value of the series is 0 which is the case for some of our values. However, the zero values correspond to times when no customers are connected, in which case we will always forecast 0 demand. Therefore, we remove these data points for the calculation of the MAPE.

In the left hand plot of Figure 3.4 we compare the MAE for all model types calculated on the training data. In general, the MAE of the DSR methods is worse than the ASM methods at all forecast lead times. Unsurprisingly the MAE gets worse at longer lead times for all methods.

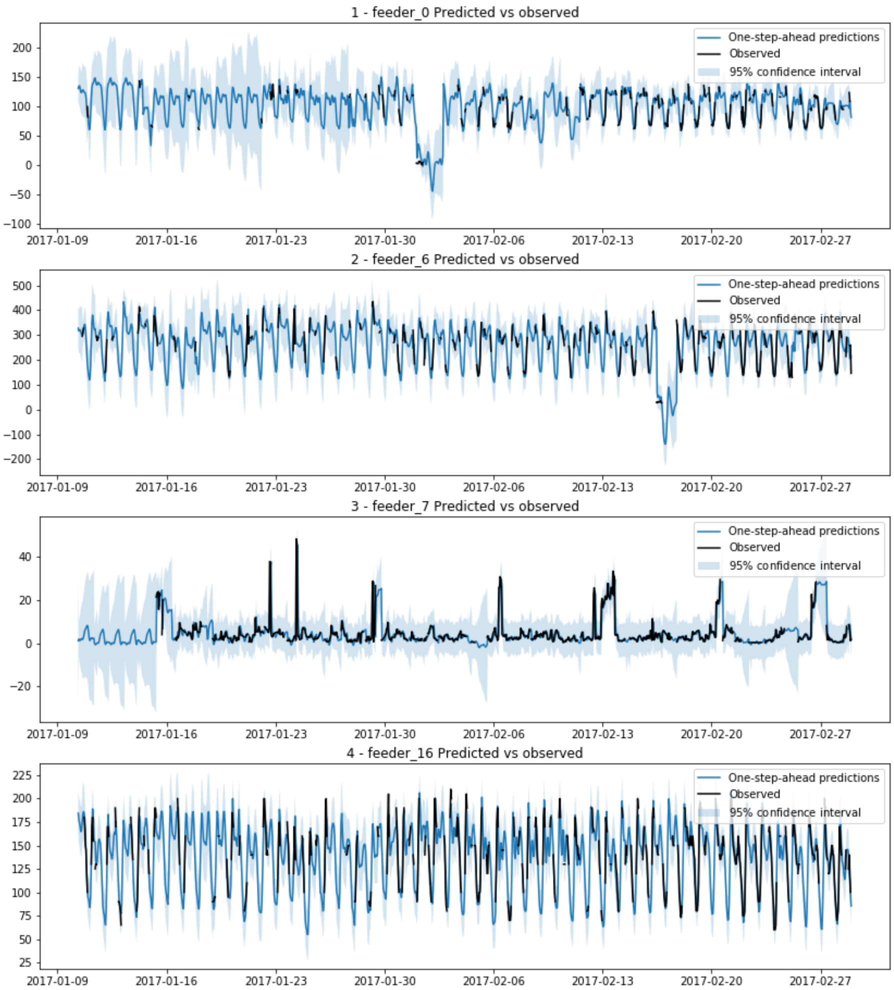


Figure 3.2: Observations compared to fitted values from structural model for a selection of feeders

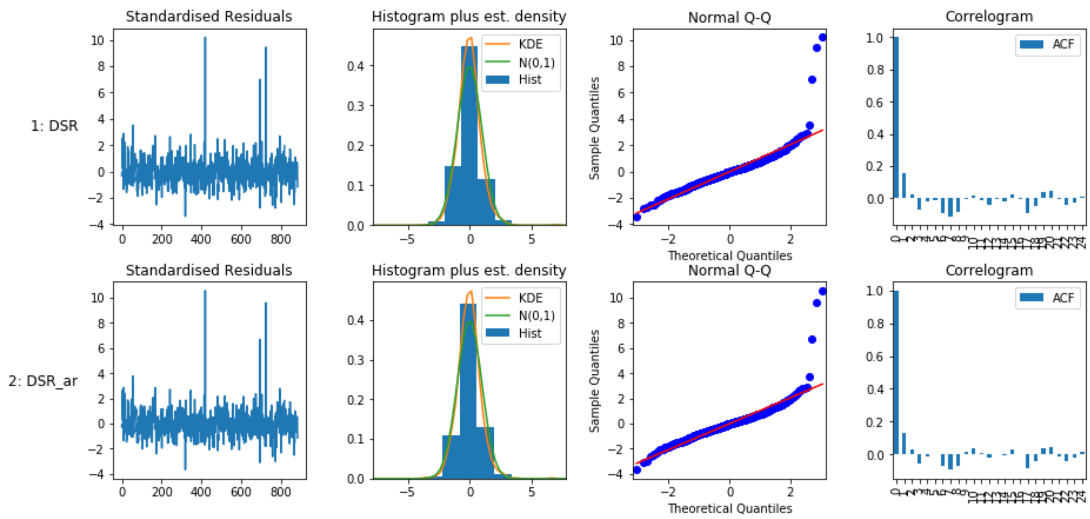


Figure 3.3: Diagnostic plots for DSR models

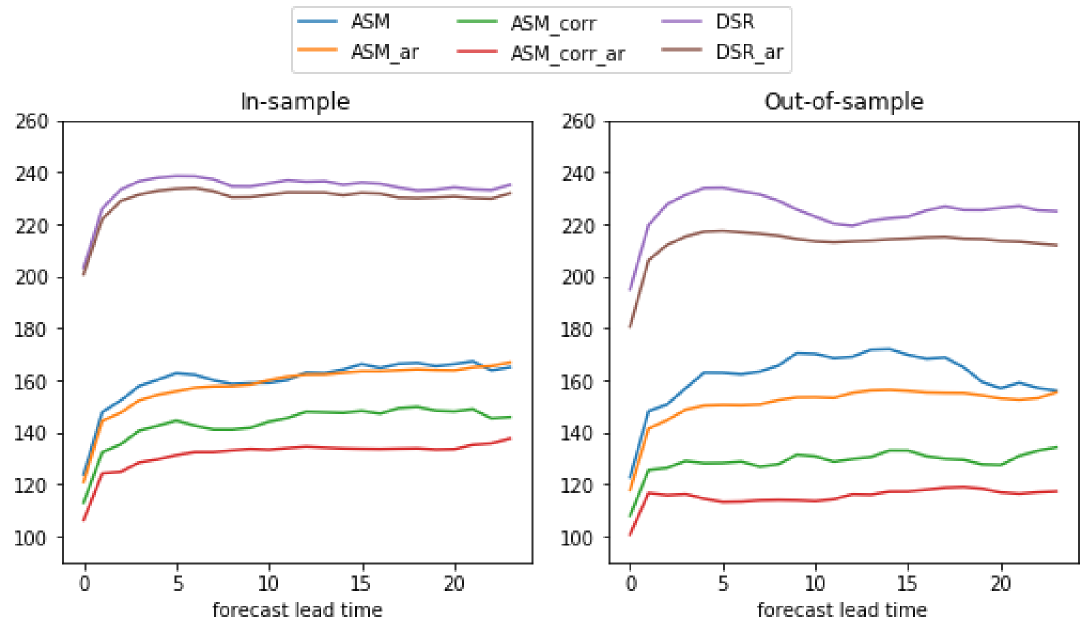


Figure 3.4: In-Sample and out-of-sample MAE

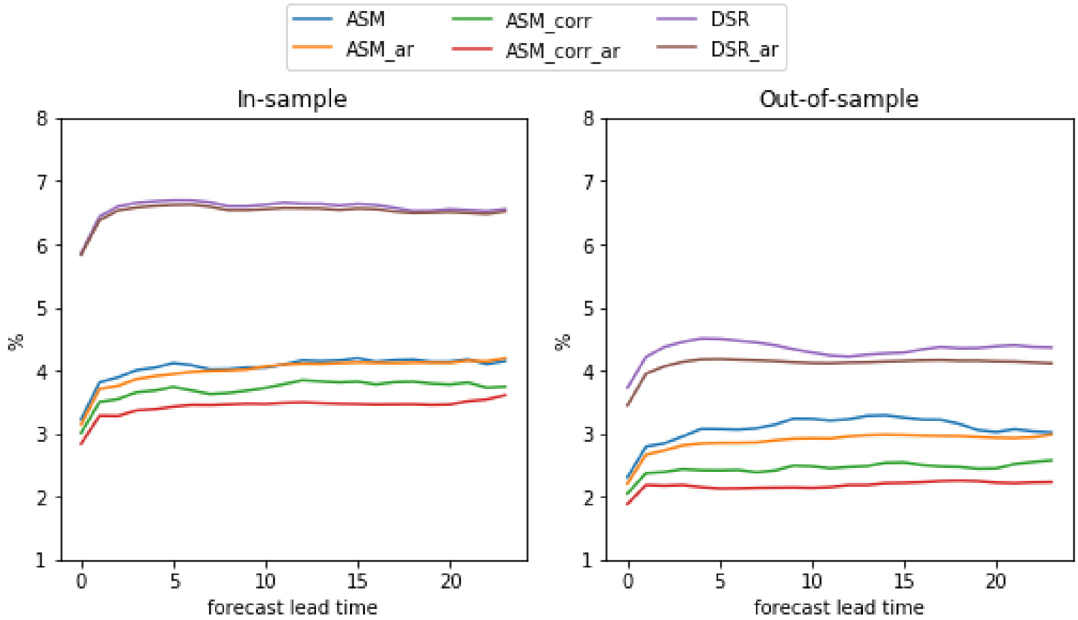


Figure 3.5: In-sample and out-of-sample MAPE

The biggest increase in MAE is between 1 and 2 periods ahead and thereafter the MAE stays at about the same level. We can also see in this plot that the forecast performance of the DSR model is not substantially improved by including AR components in the coefficients.

To get a better idea of the magnitude of the errors we can consider the MAPE. The left hand side of Figure 3.5 shows this as calculated on the training data for all models. The best performing ASM model has a MAPE of 3% at one hour ahead. This is the model supplemented with AR errors and with the DSR unsuppressed demand estimate as an exogenous covariate.

The MAE is similar on the train and test data for all our models which provides evidence that the models are not over-fitted. The MAPE improves on the test set. This is because over the test period the average load supplied increased compared to the training data, but the forecast error remained approximately constant resulting in a smaller percentage error. The lower MAPE for the training data does not therefore indicate that the forecast performance was better on the test set.

We can also see from these results that supplementing the ASM and DSR models with autoregressive terms gives an uplift in forecasting performance, as does factoring in correlations in the ASM models. This uplift in performance remains in the test set so the addition of more parameters to fit does not seem to result in over-fitting.

It is interesting that the ASM approach gives better forecasts of the suppressed load data than the DSR approach because the underlying structural models for individual feeders satisfy the modelling assumption of the LGSSM less well than the DSR model. The ASM approach is

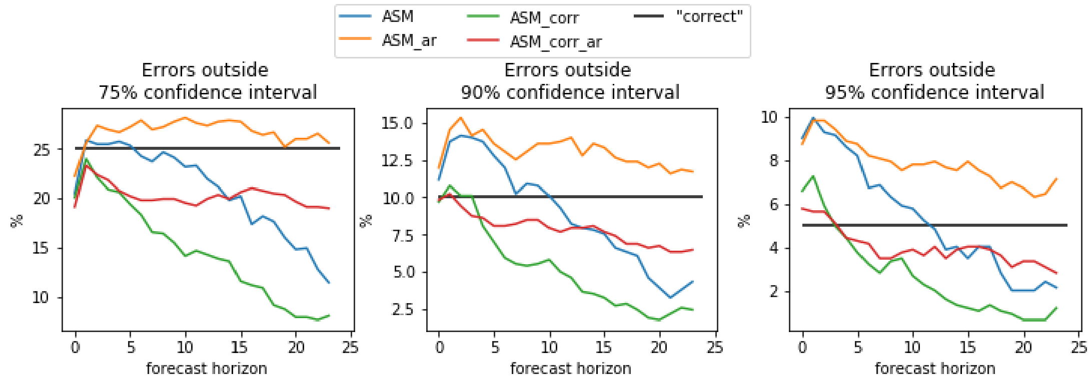


Figure 3.6: Proportion of errors from ASM approach falling outside theoretical 75%, 90% and 95% confidence intervals

able to take advantage of more information by using the demand series for each feeder.

3.4.4 Forecast Variance

It is often useful to have a probabilistic forecast of demand in addition to a point estimate. As a part of running the Kalman Filter we automatically calculate the theoretical forecast variance for all of the state space models (see Section 3.3.5). Combined with the forecast mean this gives a Gaussian probability distribution for the demand. On the assumption of independence between the forecast errors for each feeder the forecast variance of the ASM approach can be calculated by summing the variance for each feeder. We compare the theoretical forecast error variance with the empirical variance by calculating the proportion of the residuals that fall outside a theoretical 95% confidence interval.

We compare the theoretical forecast error variance with the empirical variance by calculating the proportion of the residuals in the cross validation set that fall outside a theoretical 75%, 90% and 95% confidence interval. If the distribution of the empirical variances matched the theoretical distribution of the fitted state space model then we would expect that 25%, 10% and 5% of the errors fall outside these ranges respectively. Figure 3.6 shows this at forecast lead times up to 24 for the 4 ASM models compared to a line representing the correct level. Figure 3.7 makes the same comparison for the DSR model.

Looking first at the DSR model (see Figure 3.7) we see that fewer errors fall outside the theoretical confidence interval than would be expected for all 3 confidence levels and the proportion decreases as the forecast lead time increases. This implies that the variance of the forecast errors is overestimate by the theoretical model and this effect worsens with the forecast lead time.

For the ASM and ASM_corr models (see Figure 3.6), which don't have an autoregressive component the proportion of errors falling outsize the confidence interval also declines with

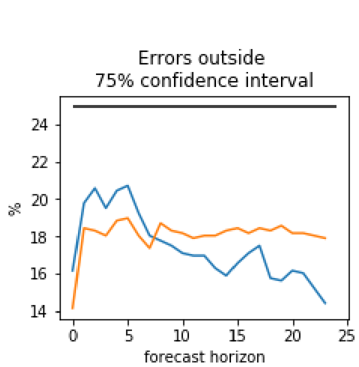


Figure 3.7: Proportion of errors from DSR approach falling outside theoretical 75%, 90% and 95% confidence intervals

the forecast lead time. At short lead times (up to 6 hours) the proportion outside the 75% confidence interval is approximately correct for the ASM model.

The DSR_ar, ASM_ar and ASM_corr_ar models are distinguished from the DSR, ASM and ASM_corr models respectively by having a state in each feeder model which follows an AR(1) process. The effect of this is that the proportion of forecast errors falling outside the theoretical confidence interval declines less quickly with the forecast lead time compared to the models without an AR term. The theoretical forecast variance from the ASM_corr_ar model gives the best approximation to the true forecast variance. It slightly underestimates the forecast variance on the test set at lead times up to 6 and slightly overestimates the variance after this.

There seem to be two effects underlying these results. Firstly, because for all the ASM models the errors are non-Gaussian then even for correct variance the confidence interval underestimates the number of extreme observations. This effect is less significant for the ASM_corr and ASM_corr_ar. These models account for the correlations between feeders which suggests that the assumption of independence between forecast errors in the ASM and ASM_ar models is not well satisfied in practice. The second effect is that the non-stationarity of the models is increasingly implausible at long lead times. Electricity demand cannot not in fact increase or decrease without limit which is allowed by these models. As a result, the theoretical forecast variance is an overestimate of the true error variance for long lead times which reduces the number of forecast errors falling outside the theoretical confidence interval.

For short forecast lead times the former effect means that the proportion of forecast errors outside the theoretical confidence interval is underestimated. However, at longer lead times the non-stationarity effect dominates, reducing the proportion of forecast errors outside the theoretical confidence interval until it is overestimated.

The ASM_corr_ar is least affected by these problems. This can be seen in Figure 3.6 because the proportion of ASM_corr_ar errors outside the theoretical 95% confidence is reasonably close to 5% at all forecast lead times. The addition of the exogenous covariate which is correlated between feeders means it is less likely to underestimate the variance. Furthermore, the non-stationarity problem is less significant because the autoregressive term is stationary and the variance of the other state disturbance terms are small relative to AR disturbance term.

3.5 Demand Estimation

We have so far evaluated our models based on their forecasting performance. These forecasts are conditional on a future policy of load shedding. That is, they require us to specify which feeders will be connected in order to provide a forecast for the resultant demand. We believe that these methods will be useful for short term operational management of the network.

We now consider our methods from the point of view of trying to estimate the historical unsuppressed demand. We observed in Section 3.1.1 that medium to long term forecasts of electricity demand are problematic without a historic dataset of unsuppressed demand. Both the methods which we have proposed imply an estimate of unsuppressed demand

Consider the ASM case. Let $\alpha_{i,t}$ be the underlying state of the time series of feeder i at time t and $Z_{i,t}$ be the design matrix such that the observed demand on feeder i is given by $Z_{i,t}\alpha_{i,t}$ plus a normally distributed error term. If the state was known then the unsuppressed demand at time t could be calculated by the implied demand from all feeders if they were connected: $\sum_i Z_{i,t}\alpha_{i,t}$

Next consider the DSR model. We estimate regression coefficients $\beta_{i,t}$ for each customer category i. The observed demand at time t is then given by the sum of each regression coefficient multiplied by the proportion customers of that category that are connected plus an observation error: $\sum_i n_{i,t}\beta_{i,t} + \varepsilon_t$. In Section 3.3.2 we described a state space form of this model where the regression coefficients are contained in the state $\alpha_{i,t}$ and the proportion of connected customers is in the design matrix $Z_{i,t}$. The unsuppressed demand is given by taking $n_{i,t} = 1$ for every customer category.

In both cases the unsuppressed demand can be calculated using an estimate of the state. We previously used the Kalman filter to estimate the state at time t conditional on the observations prior to t. For a post-hoc estimation of demand we would like to estimate the state conditional on all the data before and after the time period of interest. The *Kalman Smoother* can be used to estimate α_t given the entire history of a time series $y = y_1, \ldots, y_n$, i.e. $\hat{\alpha}_t = \mathbb{E}[\alpha_t | y]$.

The Kalman Smoother consists of a recursion that is applied in reverse order to the data:

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t, \tag{3.8}$$
$$\hat{\alpha}_t = a_t P_t r_{t-1}, \tag{3.9}$$
$$r_0 = 0 \tag{3.10}$$

where F_t, L_t, P_t, v_t and a_t are given by applying the Kalman Filtering recursions to the data.

Using this method, we estimated the unsuppressed load on the 76 feeders used in the forecasting evaluation during the period that was used to fit the model parameters. Of the 6 candidate models considered in Section 3.4.1 we compare ASM_ar, the ASM model with an autoregressive error, and DSR_ar, the DSR model with autoregressive errors for each coefficient.

The load-duration curve is calculated by sorting the estimated load in descending order as shown in Figure 3.8.

It is reassuring that there is a good correspondence between the load duration curves calculated by two different methods. However, we note that there is a pronounced divergence between the two estimates at the higher end of the load duration curve.

Estimating the peak demand is particularly important in power system planning because the system capacity should be designed with this in mind. However, it should be noted that in the context of a power system that is starting from a point of capacity inadequacy, this will surely be added in an incremental way. If capacity grows relative to demand and the frequency of load shedding decreases, then more accurate estimates of peak demand will be possible.

We suggest two possible reasons for this divergence. At the left-hand end of the load duration curve the largest divergence between the model is accounted for by a very small number of values. Several of these very high values coincide with very large forecast errors that can best be explained by outlying values due to non-Gaussian observation noise. In Figure 3.9 we plot the standardized forecast errors over time. The black lines represent the 0.00001% critical values from the standard normal distribution. Without the need for formal analysis we can see that these observations are very unlikely according to the assumptions of the model. Excluding these 4 observations trims the left-hand end of the load duration curve slightly. However, it cannot account for the systematic difference between the DSR and ASM results.

The ability to handle missing values is a benefit of the state space model framework and the Kalman filter. However, in our case data is not missing completely at random but is more likely to be missing at peak hours. It is possible that this may cause the estimated models to perform worse at peak hours because there is less data to fit the models to at these hours.

To test this we simulated an autoregressive process and added this to a periodic deterministic mean. We then randomly removed values in such a way that values are more likely to be

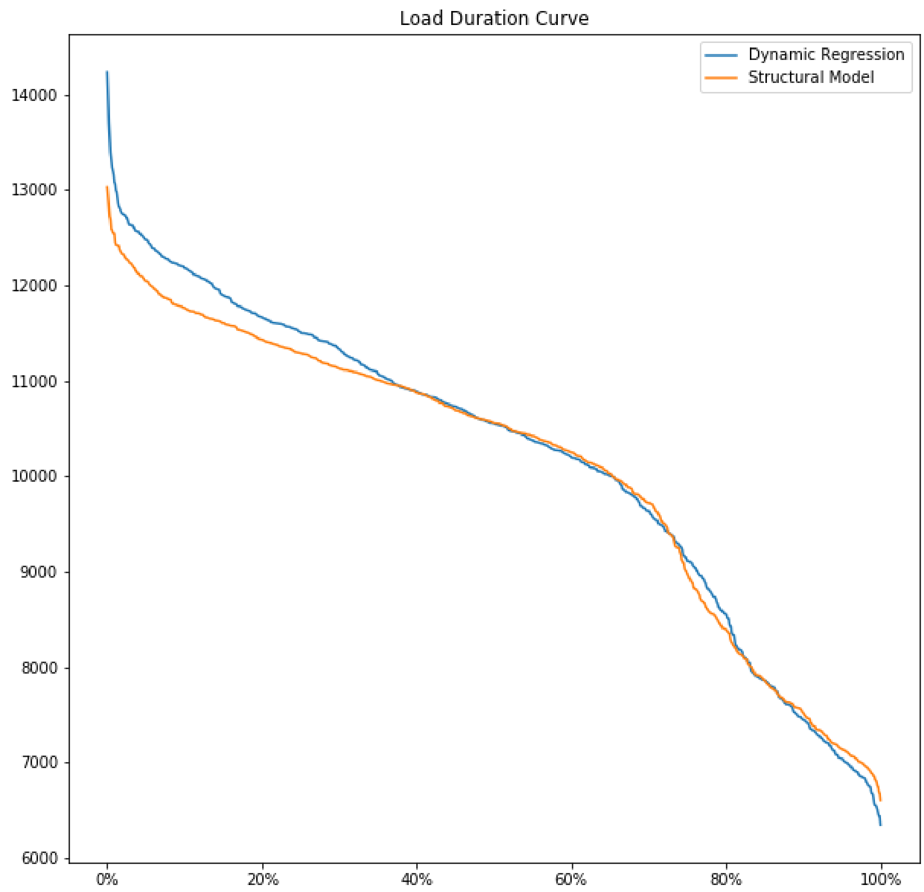


Figure 3.8: Estimated load-duration curve according to DSR and ASM model

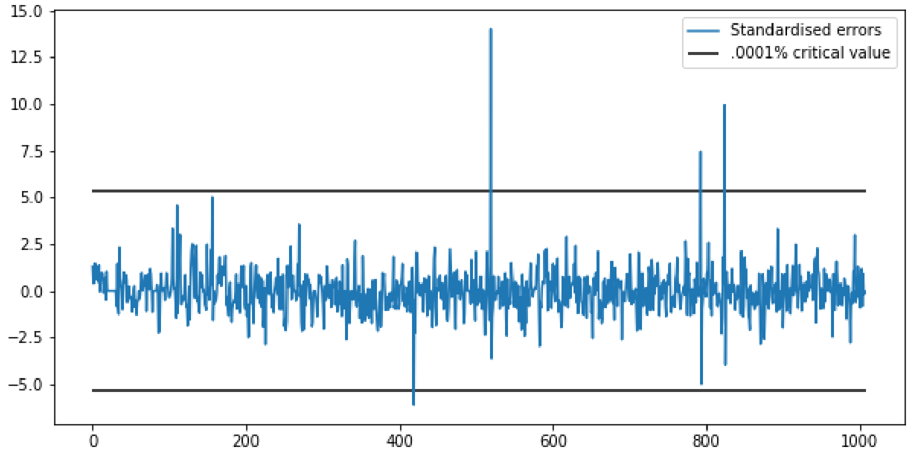


Figure 3.9: Standardised residuals from DSM model compared to the .00001% critical values of the standard normal distribution

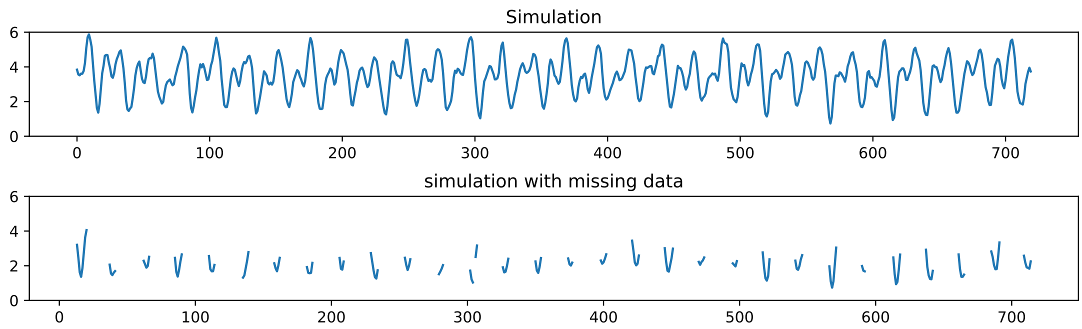


Figure 3.10: Simulated feeder data

removed where the mean value is high. This simulates the effect of load shedding which is more common in peak hours. In order to do this we used a function which takes the size of a value, normalized by the range of values in the sample, and maps this to a probability of removal using the logistic function shown in Figure 3.11. The largest values were removed with a probability close to 1 while small values are very unlikely to be missing. A illustrative subset of the simulated series and the simulated series after data has been removed is shown in Figure 3.10.

We then fitted a structural time series model to this data set. The smoothed estimated values are shown in Figure 3.12 along with the actual values in the sample. It can be seen that there is a slight tendency to underestimate the peaks of the series. When the values are re-ordered in the style of a load duration curve as in Figure 3.13 we can see this has a similar effect to that observed in the actual load duration curve.

We repeated this experiment with a larger amount of data by using a sample length of 300 days,

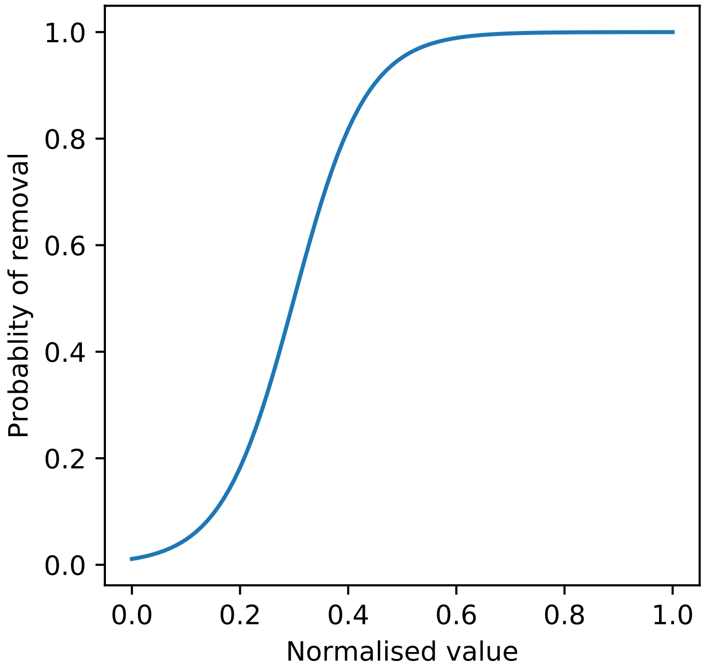


Figure 3.11: Logistic function to determine probability of removal

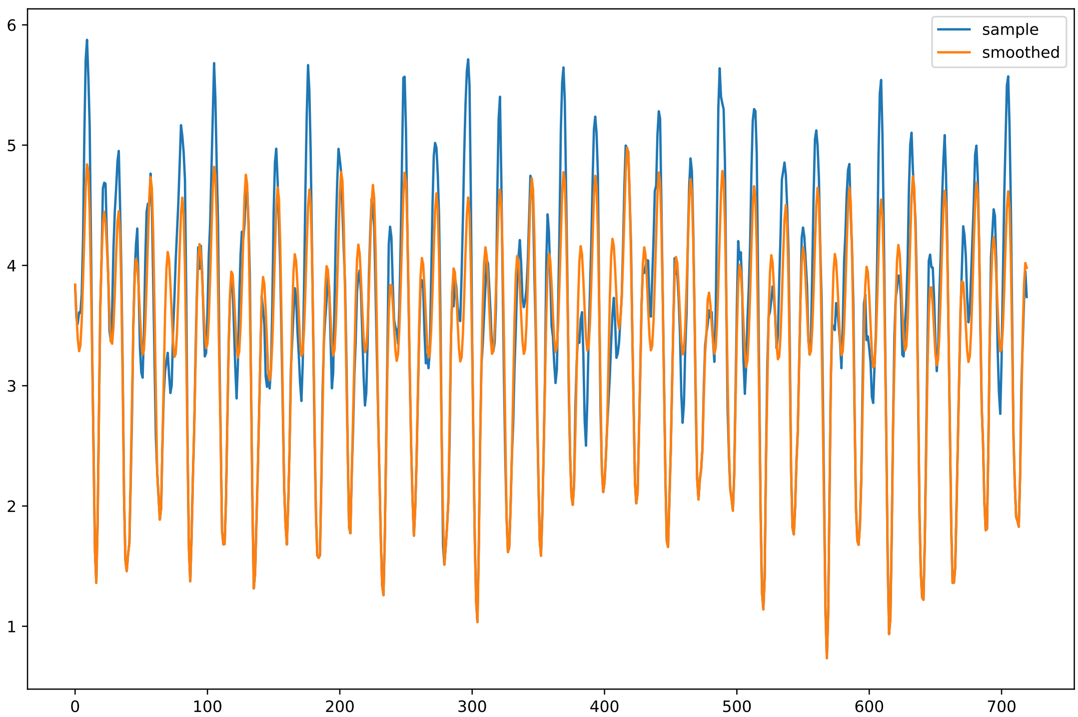


Figure 3.12: Comparison of simulated data to smoothed values from a structural model

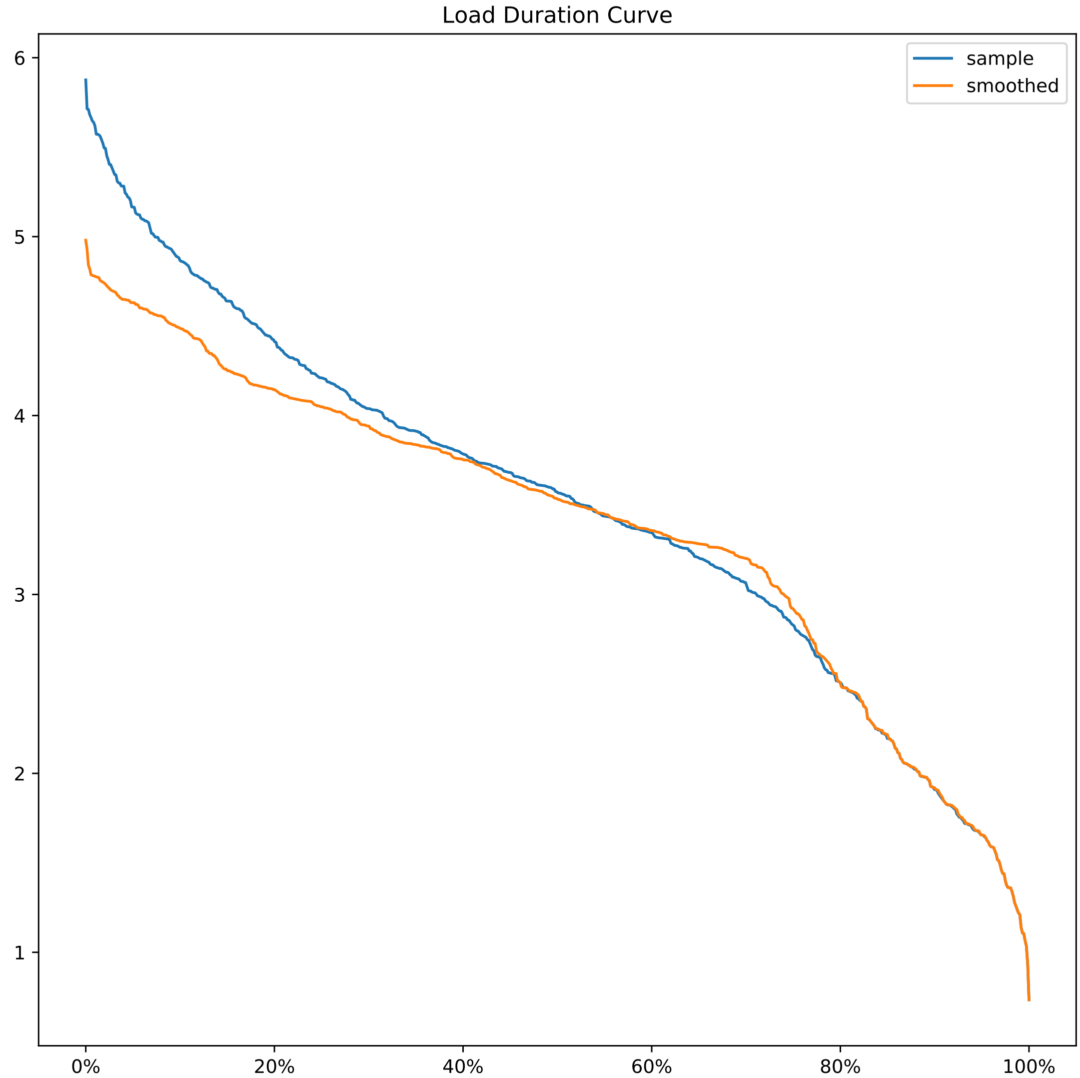


Figure 3.13: Load duration curve for 720 points of simulated data compared to load duration curve of smoothed values from structural model estimated with missing data

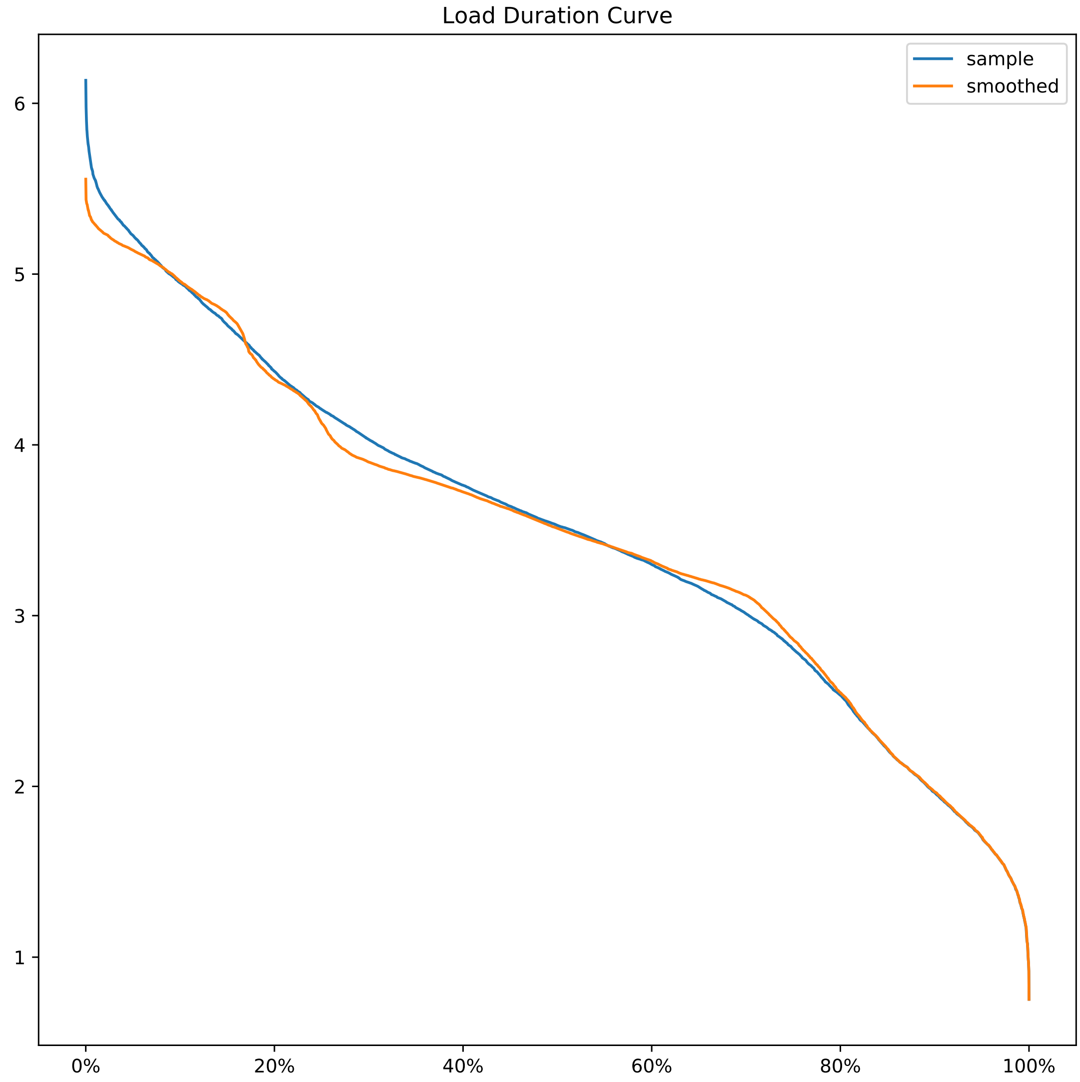


Figure 3.14: Load duration curve of 12,000 points of simulated data compared to load duration curve of smoothed values from structural model estimated with missing data

or 12,000 hours. The estimated load duration curve for this sample is much better, as shown in Figure 3.14. This suggests that the ASM model could accurately model the peaks with a sample that is sufficiently long to have enough connections during peak hours. However, simply using a longer sample raises some other issues when applied to real data because of the likelihood of structural breaks existing in a longer dataset. A structural break is a change in the probabilistic structure of the stochastic process that could be caused by things like “unusual” days such as holidays, total system collapses or connection of more customers.

Finally, another possibly reason for the discrepancy is that the ASM model does not account for correlations between demand on different feeders. The highest demand days occur when demand in the DisCo is above average. This may cause the ASM method to underestimate the unsuppressed demand because there are no observations from the disconnected feeders from which to infer that they are also above average. The same is also likely to be true for low demand hours, but any overestimation of demand is likely to be less pronounced because low demand is likely to mean that more feeders are connected, improving the estimate from the ASM method. There does seem to be a very slight effect that the load duration curve from the ASM method is higher than the DSR method at the lowest demand days.

3.6 Further Improvements to the Methods Presented

More analysis is needed to determine if the violations of the linear Gaussian assumptions of the structural models for distribution feeders is intrinsic to the nature of electricity demand on these feeders. Understanding this requires other possible causes for the violations of the model assumptions to be systematically eliminated. Unfortunately, this was not something we were able to carry out because more data is required.

The first step would be to eliminate, as far as possible, anomalies in the data that are due to poor data entry. We deliberately refrained from carrying out substantial "cleaning" of the data in order to see if we could design models that are sufficiently robust given existing data recording methods. By using automated systems or validating data at the point of recording the frequency of erroneous data could probably be reduced.

We also have not considered how partial disconnections might affect the observed demand. By partial disconnections we mean equipment failures resulting in a proportion of the customers connected to a feeder being disconnected. This would lead to a lower level of demand being reported on a feeder than would happen otherwise and would manifest itself in the models by structural breaks and time dependent variance. The importance of this effect is difficult to estimate without knowing when such events have occurred. If this information was available, we could compare the fitted models to see if the model assumptions are better satisfied for feeders where partial disconnections do not occur. We could also compensate for these events by using dummy "intervention variables" that allow an arbitrary "jump" in the data at these points. Alternatively, we could simply fit separate models between partial disconnection events. We would expect this to result in better fitting models because the parameter estimation will not be biased by having to account for large residual errors at the time steps where disconnection events occur.

Incorporating a model for weekly seasonality and adding regression effects to account for changes in demand subsequent to periods of load shedding might improve forecasting performance. However, the resulting increase in states and parameters means that we may need to fit the models to longer samples to avoid over-fitting. We saw in Section 3.5 that we may have had some difficulty estimating the true peak demand using a structural model as a result of not using a long enough sample. This would probably introduce more "unusual" days into the sample like holidays or total system collapses that we would need to account for (possibly with dummy intervention variables). We also believe that dealing with the high frequency of data anomalies is a prerequisite for reliable modelling of subtle effects (relative to daily seasonality).

We have some reason to believe that these steps would result in finding models that better match the linear Gaussian assumptions of the structural model. Consider for example row 4 of the diagnostic plots in Figure 3.1. The forecast residuals of this model are well described by a normal distribution and appear to have a stationary mean and variance. The fact that at least some of the feeders are well described by a linear Gaussian model suggests that other feeders could be as well if we could account for certain unusual features in the data along the lines suggested above.

However, if it were concluded that the limitations of the model are due to intrinsic features of the data generating process then substantial improvements in forecasting performance or signal extraction will require a change in the model used for the individual feeders. Two time series modelling approaches that are appealing to deal with structural breaks, outliers and dynamic changes in variance parameters are: firstly, approximate filtering or importance sampling methods for non-Gaussian state space models as developed in [17]; and secondly, Generalized Autoregressive Score (GAS), also known as Dynamic Conditional Score (DCS) models such as those developed in [11] and [25].

3.7 Conclusions

In this chapter we have formulated two methods to forecast the suppressed electricity demand in an electricity system subject to constant load shedding. Our ASM approach models each feeder demand time series as a structural model with a random walk component and 24-hour seasonality. We then aggregate these forecasts by summing the forecasts for the connected feeders. The DSR method models the unsuppressed demand as a linear regression where the proportions of various types of customers that are connected are the independent variables. We posit that these regression coefficients evolve according to the dynamics a structural time series model and use the Kalman filter estimate these coefficients conditional on the observed data.

We considered some extensions to the basic ASM method: 1) We supplemented each structural model with a stationary autoregressive component (denoted ASM_ar); 2) we used the implied estimate of the unsuppressed demand from the DSR model as an exogenous regression variable (denoted ASM_corr); and 3) we considered both of these features together (denoted ASM_corr_ar). We also formulated a version of the DSR model in which the regression coefficients have a stationary autoregressive component (denoted ASM_ar).

In general the ASM type models gave the best forecasts according to the MAPE and MAE.

The ASM_corr_ar model had the best overall performance with a MAPE of 3% for the one-step-ahead forecasts rising to 3.5% by 24-steps-ahead. In addition to improving the forecast performance of the model, incorporating a stationary autoregressive component and exogenous regressor into the ASM_corr_ar model improved the estimate of the forecast variance.

We noted that the underlying model assumptions of the LGSSM seem to be poorly satisfied for the structural models in the ASM method. For many of the feeders we modelled there is evidence of non-stationary and non-Gaussian standardized forecast errors. In spite of this, the ASM models are able to provide accurate forecasts. [48] shows that the Kalman filter is the optimal *linear* filter even for non-Gaussian noise which means that, in terms of point predictions the Kalman filter is reliable even when the measurement noise is non-Gaussian. Hence the ASM model can provide better forecasts than the DSR model even though the model assumptions of the LGSS model are not as well satisfied by the ASM model as by the DSR model. However, this compromises the probability distributions derived from the model.

The DSR method has the advantage that it can be applied to feeders where demand data has not been recorded. This is an advantage in the context of power systems in developing countries where data recording and communication procedures and infrastructure are often poor. It is also possible that the DSR method could be improved by supplementing that dataset with other variables. For example, variations in the number of illegal connections on each feeder could be included in the data if better customer enumeration was carried out. Socio-economic data could also be incorporated into the model to account for non-homogeneity of customers on each feeder. Utilities that want to use this approach should therefore focus on obtaining accurate and detailed information about the population supplied by each feeder.

Despite some limitations, our models can provide accurate forecasts of the load conditional on a given load shedding policy. We have reason to believe that with further development the quality of the estimated distribution of the forecast errors could be improved. In Chapter 4 we will investigate whether our forecasts can effectively be used to provide decision support to DSOs.

Chapter 4

Distribution Level Load Shedding

4.1 Introduction

A long-term power shortage necessitates a system of demand management to balance electrical supply and load. The SO's direct control of loads is limited to disconnecting large areas at the interface between the transmission and distribution systems. Disconnection of regions of the network at this level may be necessary in extremis to prevent system collapse in the case of the unexpected loss of a generator or transmission line. However, most load shedding carried out on a routine basis is undertaken by DSOs. The DSOs can control the power drawn by connecting or disconnecting distribution feeders. The role of the SO is to instruct the DSOs how much electricity they should consume.

A variety of considerations affect the decisions DSOs make to respond to the SO's instructions. These may include: maximizing utilization of available energy; avoiding overconsumption of energy relative to the SO's instructions; providing a predictable, if intermittent, power supply to customers; meeting service quality standards such as a minimum connection frequency or maximum outage length; and maximizing revenue by directing energy to feeders with lower losses/higher average tariffs. This is a difficult problem because the underlying demand process is stochastic and is also imperfectly observable because of regular disconnection of feeders.

In the Nigerian case the Distribution Code makes provisions for a variety of methods of demand management including voltage reduction, automatic under-frequency load shedding and manual load shedding:

> Distribution Code Section 3.5.3
> Where instructed by the SO, temporary load shedding shall be carried out to maintain the load generation balance. This may also be necessary due to lack of generation, loss of any circuit, equipment or any other operational contingency.

To the extent that demand management through voltage reduction is possible, the long run nature of the problem means that consumers are likely to have compensated (by installing more appliances, running them at higher settings etc.). In addition to these categories the distribution code obliges DisCos to respond to the SOs instructions to maintain load generation balance by shedding load manually. In practice this is a continual necessity. The Distribution Code also stipulates that DisCos should notify affected customers in advance:

> Distribution Code Section 3.5.2
> In the event of load shedding under the DisCo's planned load shedding rotas, the public shall be promptly notified of such arrangements through the media or on a web site.

This often does not happen because the DisCos lack reliable methods for forecasting demand and systematically scheduling load shedding in advance. At present these objectives are managed in a distributed and ad-hoc way by decisions of local substation operators who have tacit knowledge of the typical loading on feeders.

The market regulations also provide for a system of penalties to distribution companies that consume more than their allocated energy, although it is unclear to what extent these are enforced in practice.

To recapitulate, DSOs face a continual problem of selecting which feeders to connect because the latent electricity demand would always exceed the available supply. Power flows to the consumers on a connected feeder according to the level of latent demand on that feeder. We call the realized demand on the set of all connected feeders the *suppressed demand*. Demand evolves stochastically, so the future suppressed demand and resulting energy sales cannot be known with certainty. Selling energy is the privately owned DSO's primary objective so it is important to maximize revenue by connecting as many customers as possible. In addition, because of varying customer tariff rates and collection losses the revenue received as a result of supplying energy varies by feeder. As a result, the DSO should prioritize certain feeders in order to maximize revenues.

However, in addition to revenue maximization several other considerations are important:

Firstly, DSOs receive instructions from the SO to limit their suppressed demand to a given level. Following the Nigerian convention, we call this the DSO's *load allocation*. If suppressed demand exceeds the load allocation, then this will contribute to network problems which may result in the SO disconnecting parts of the DSO's network and in extreme cases could lead to a system collapse. The DSO could manage this risk with a chance constrained approach, that is, maximizing their expected revenue from energy sales subject to an upper limit on the probability of suppressed demand exceeding the load allocation. Alternatively, if the regulatory regime imposes penalties on DSOs that exceed their load allocation, as in Nigeria, the DSO may prefer to find a policy that maximizes expected revenue minus penalties from overconsumption.

Secondly, because of regulatory obligations or commercial strategy the DSO may also wish to satisfy certain supply quality constraints for all feeders. These might include limits to the minimum connection frequency per day or the maximum length of supply interruptions.

It is also possible that the physical limits of the DSO's electricity distribution network may mean that certain combinations of connected feeders are infeasible. These limits include power flow limits on transformers and distribution lines as well as bus voltage limits. However, if the distribution network is appropriately sized and the main factor constraining supply is the generation and transmission system it may not be necessary to consider technical constraints on the distribution network.

We propose to model this as a discrete time problem of selecting which feeders to connect in blocks of one hour in length. We suggest that the DSO's problem can be captured in an optimization framework that resembles a stochastic variant of the classic "knapsack problem". This allows us to prioritize feeders based on their average tariff and level of non-technical losses while managing the risk over over-consumption. We will show how supply quality constraints can be incorporated into this problem.

Of course, it is not strictly true that the DSO has to commit to keeping feeders on for a whole hour. One might wonder if the DSO can do better by connecting feeders for a very short period in order to test the level of demand on that feeder. This is unlikely to work in practice because of the phenomena of cold load pickup. In the first 15 seconds of reconnection current magnitude can be 10 to 15 times the normal value due to the flow of current to cold lamp filaments, re-magnetization of distribution transformers and starting current needed for the acceleration of a motor [37]. A large increase in demand can also be caused because of thermostatically controlled devices which, if they have not been powered by backup generation, will all switch to the 'on' state at once. On the other hand, a study of several days of rotating power interruptions in Taiwan [38] found that in the first minutes of reconnection loads were below their expected level and increased to about 80% of their expected steady state value in the first 4 minutes of reconnection. They suggest this is because many devices must be manually restarted. This means that the load observed on a feeder in the first minutes of reconnection is not a reliable guide to the demand on that feeder. Therefore, models are needed that predict the demand on a feeder after it has been connected for enough time for the cold-load pickup effect to subside. Methods are required that can plan based on these models.

In order to analyze this problem, we now introduce the *Stochastic Knapsack Problem*.

4.2 The Stochastic Knapsack Problem with Random Weights

The standard Knapsack problem is defined as follows: given a set of objects with a *value* and a *weight*, choose a subset of the objects to maximize the sum of the values subject to an upper limit on the sum of weights. The knapsack problem is NP hard but can be solved in pseudo-polynomial time by dynamic programming [23].

The knapsack problem has many practical applications including cutting problems and resource allocation problems[40]. It is also frequently found as a sub-problem in large scale integer optimization techniques such as column generation for the generalized assignment problem or the bandwidth packing problem [23].

In stochastic variants of the problem one or more of the parameters is a random variable. Various practical applications of stochastic knapsack problems have been proposed in the literature including: purchase of financial products with a limited budget [36]; assigning uncertain demands to production capacity[40]; target market selection [50]; selecting from multiple contracts with uncertain resource requirements or an electrical supplier who can supply power to groups of customers [35]

We will focus on the *stochastic knapsack problem with random weights* because this is analogous to our Feeder Scheduling problem. This differs from the standard knapsack problem because the item weights are random variables and a linear penalty term is imposed for any violation of the knapsack weight constraint when the values of the item weights are realized.

In the feeder scheduling problem, the knapsack capacity corresponds to the SO's instruction to limit consumption to a given level. The penalty per unit weight in excess of this corresponds to the penalty for overconsumption of power. The "weight" of each item corresponds to the demand on each feeder. The reward per unit weight corresponds to the expected revenue from power supply to a feeder.

[40] shows that a simple way of incorporating stochastic knapsack capacity is to include an additional item in the problem with mean weight 0, variance equal to the variance of the knapsack capacity and to fix the decision variable for this item so that it must be included in the knapsack. We assumed above that the DSO's threshold for overconsumption penalties is fixed once the DSO receives its load allocation. We could also consider the case where the DSO's load allocation for each hour may be subject to random changes using this approach.

In our case the problem must be solved repeatedly. At every stage the distribution of the demand on each feeder reflects our state of knowledge about that feeder. This will depend on the past pattern of connections and disconnections on that feeder. In this section we will discuss the standard single shot knapsack problem for ease of exposition. We extend our notation to the multi-period case in the next section.

We now develop a notation to describe the problem based on [40].

4.2.1 Notation for the Stochastic Knapsack Problem

Suppose we have n items to choose from. Denote the item weights by the random vector $d = (D_1, ..., D_n)$. The decision to take item i or not is modelled by the binary variable x_i. The total weight of the selected items will be denoted $D(\boldsymbol{x}) = \sum_{i=1}^{n} D_i x_i$.

Let us define $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)^T$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, ... \sigma_n^2)^T$ and $\boldsymbol{x} = (x_1, ..., x_n)^T$. For independently normally distributed item weights we have $D(\boldsymbol{x}) \sim N(\boldsymbol{\mu}^T \boldsymbol{x}, \boldsymbol{\sigma}^{2^T} \boldsymbol{x}^2)$.

Let r_i denote the reward per unit weight if item i is selected and $\boldsymbol{r} = (r_1, ..., r_n)^T$. The vector of expected rewards for each item (if selected) is therefore given by $(\boldsymbol{r} \circ \boldsymbol{\mu})$, that is the element-wise multiplication of $\boldsymbol{r}$ and $\boldsymbol{\mu}$. Let p denote the penalty per unit of weight in excess of the knapsack capacity, denoted C.

The expected value of choosing the items according to $\boldsymbol{x}$ is therefore:

$$SKP(\boldsymbol{x}) = (\boldsymbol{r} \circ \boldsymbol{\mu})^T \boldsymbol{x} - p\mathbb{E}[(D(\boldsymbol{x}) - C)^+]$$

The expected penalty is proportional to expected excess demand $\mathbb{E}[(D(\boldsymbol{x}) - C)^+]$, where $(D(\boldsymbol{x}) - C)^+ = \max\{0, D(\boldsymbol{x}) - C\}$. In order to analyze this, let us assume that the average demand level is μ and therefore the average "displacement" of the realized demand from the capacity threshold is $\hat{\mu} = C - \mu$. Let us assume for the time-being that the demand has $\sigma^2 = 1$ so we can work in terms of the standard normal distribution with pdf denoted by $\phi(z)$. The expected penalty is therefore proportional to $\int_{\hat{\mu}}^{\infty}(z - \hat{\mu})\phi(z)dz$, i.e. the expected excess demand over the capacity. Note that this can be readily evaluated for known $\hat{\mu}$ since this is the standard normal loss function:

$$L(z) = \int_{z}^{\infty}(z' - z)\phi(z')dz' = \phi(z) - z\bar{\Phi}(z) \tag{4.1}$$

where ϕ is the normal PDF and Φ is the normal CDF and $\bar{\Phi}(z) = 1 - \Phi(z)$

Now suppose that the demand has variance σ^2. We can express the expected excess demand in terms of the mean and standard deviation:

$$\hat{L}(\hat{\mu}, \sigma) = \int_{\hat{\mu}}^{\infty}(z - \hat{\mu}) f(z|0, \sigma^2)dz \tag{4.2}$$

where $f(z|0, \sigma^2)$ is the pdf of a normal distribution with mean 0 and variance σ^2. $f(z|0, \sigma^2)$ can be expressed in terms of the standard normal pdf, i.e. $f(z|0, \sigma^2) = 1/\sigma\phi(z/\sigma)$. Now let $\hat{z} = z/\sigma$, which implies that $dz = \sigma d\hat{z}$, thus:

$$\hat{L}(\hat{\mu}, \sigma) = \int_{\hat{\mu}/\sigma}^{\infty}(\sigma\hat{z} - \hat{\mu})\phi(\hat{z})d\hat{z}$$

Taking out a factor of σ:

$$\hat{L}(\hat{\mu}, \sigma) = -\sigma \int_{\hat{\mu}/\sigma}^{\infty}(\hat{z} - \frac{\hat{\mu}}{\sigma})\phi(\hat{z})d\hat{z}$$

Therefore

$$\hat{L}(\hat{\mu}, \sigma) = \sigma L(\frac{\hat{\mu}}{\sigma}) \tag{4.3}$$

This means that the expected excess demand when the threshold is a fixed number of standard deviations from the mean demand is proportional to σ.

Hence the stochastic knapsack problem with random weights can be reformulated as:

$$\max_{\boldsymbol{x}\in\{0,1\}^n} SKP(\boldsymbol{x})$$

where:

$$SKP(\boldsymbol{x}) = (\boldsymbol{r}\circ\boldsymbol{\mu})^T\boldsymbol{x} - pL\left(\frac{C-\boldsymbol{\mu}^T\boldsymbol{x}}{\sqrt{\boldsymbol{\sigma}^{2T}\boldsymbol{x}^2}}\right)\sqrt{\boldsymbol{\sigma}^{2T}\boldsymbol{x}^2}$$

4.2.2 Reformulating the MINLP Problem to a More Tractable Problem

The stochastic knapsack problem with random weights is a Mixed Integer Non-Linear Program (MINLP). We propose to model SKP using a piecewise linearization of the non-linear terms. The two non-linear terms are the expected penalties and the standard deviation of demand.

Piecewise Linear Formulation of the Expected Penalties

We can formulate a piecewise linear approximation to the expected penalties in terms of how many multiples of the standard deviation that the mean demand is from the capacity threshold.

$L(m)$ is the expected excess demand for a normal random variable with standard deviation 1 and mean m *less than* the threshold. i.e. $C-\mu = m$ where μ is the mean demand and C is the capacity threshold. Equation 4.3 shows that if the standard deviation is σ and the mean is $m\sigma$ then the expected excess demand is $\sigma L(m)$. Suppose we have calculated $l_i = L(m_i)$ for n points $m_0,\ldots,m_n$ and $\hat{\lambda}_0,\ldots,\hat{\lambda}_n$ are associated weight variables, the convex hull of a piecewise linear approximation to $\hat{L}(\sigma,\hat{\mu})$ is defined by

$\hat{\mu} = \sigma\sum\hat{\lambda}_i m_i$

$\hat{L} = \sigma\sum\hat{\lambda}_i l_i$

$\hat{\lambda} = \geq 0$

$\sum\hat{\lambda}_i = 1$

σ is a variable, but we can obtain a linear formulation by taking $\lambda_i = \sigma\hat{\lambda}_i$, i.e.

$\hat{\mu} = \sum\lambda_i m_i$

$L = \sum\lambda_i l_i$

$\lambda = \geq 0$

$\sum\lambda_i = \sigma$

Piecewise Linear Formulation of Standard Deviation

The standard deviation of total demand is the square root of the sum of the variance of the selected feeders.

$$\sigma = \sqrt{\boldsymbol{\sigma}_t^{2^T} \boldsymbol{x}_t^2}$$

We can relax this to obtain a norm constraint:

$$\sigma \geq \sqrt{\boldsymbol{\sigma}_t^{2^T} \boldsymbol{x}_t^2}$$

Assuming we have linearized the expected penalties this is the only non-linear term so we could use solvers for conic programming to solve the continuous relaxation.

However, our approach is to note that $x_{i,t} \in \{0,1\}^n$, so $x_{i,t} = x_{i,t}^2$

We can therefore define the variance of the total demand as $v_t = \sum_{i=0}^{n} x_{i,t}\sigma_{i,t}^2$.

This is a linear expression for the variance in the continuous relaxation of the problem given by $\mathbf{x_t} \in [0,1]^n$. The relaxation of $\sigma = \sqrt{v}$ to $\sigma \geq \sqrt{v}$ is non-convex. We can approximate this using a piecewise linear function with the additional constraint that only two adjacent weight variables are permitted to be non-zero. In Section 4.4 we show how this can be formulated with SOS2 sets. Solvers for Mixed Integer Linear Programs (MILPs) can exploit this to solve the problem efficiently.

Comparison of the MINLP Formulation to the MILP Approximation

We show in the appendices that the continuous relaxation of $SKP(x)$ has a concave continuous relaxation. Since we are maximising this function, MINLP solvers can, in principle, solve the problem to proven optimality. However, MINLP solvers are slow relative to solvers for Mixed Integer Linear Programs or even Mixed Integer Quadratically Constrained Programs. In order to determine the practicality of solving SKP directly compared to solving a linear approximation we carried out some numerical experiments with randomly generated problem data.

We considered 10 problem instances with randomly generated the means and variances and average reward per unit. The number of items ranged from 20 to 200 in increments of 20. We scaled the problem data so that the sum of weight was the same in all problem instances.

To solve the exact MINLP formulation we used the commercial solver KNITRO. The formulation of the MILP approximation is fully described in Section 4.4. We considered formulations with 4, 8, 16, 32, 64, 128 and 256 linear pieces to see how this would affect the solution time and accuracy of approximation.

In Table 4.1 below we give the solution times for the MINLP problem and MILP approximations of each number of linear pieces. It can be seen that the solution time for the MINLP

	MILP with x linear pieces							MINLP
	4	8	16	32	64	128	256	
n_items								
20	0.64	1.95	0.16	2.02	2.06	2.14	2.13	0.33
40	0.68	2.17	0.20	1.49	2.17	1.87	0.20	0.73
60	0.31	1.99	1.94	1.87	0.32	0.30	0.27	0.47
80	2.02	1.86	1.97	2.08	2.46	1.50	1.69	27.69
100	2.13	2.23	1.60	1.73	1.45	2.03	1.38	3.91
120	2.11	2.09	2.28	1.86	1.92	1.35	1.92	6.00
140	1.05	1.88	1.44	2.18	2.04	1.42	1.38	174.23
160	1.34	1.85	1.63	2.17	1.86	0.93	1.87	275.76
180	1.22	1.87	2.01	0.98	1.64	1.38	1.62	691.54
200	0.67	0.97	1.99	0.81	2.01	1.19	1.62	617.21

Table 4.1: Solution time in seconds to *SKP* with randomly generated problem instances with 20-200 items.

problem scales poorly with the number of items. Even worse performance would be expected for the multi-period cases we will consider late in this chapter. We also note that the number of linear pieces does not strongly affect the solution time. We found that with more than 32 pieces all problems could be solved to less than 0.1% of the true solution. We therefore choose to use the MINLP approximation for our subsequent analysis.

4.3 The Feeder Scheduling Problem as a Multi-Period Stochastic Knapsack Problem

The feeder scheduling problem can be formulated as a sequence of stochastic knapsack problems where the item weights evolve according to a stochastic process.

Which feeders we decide to connect affects the future distribution of demand on each feeder because it changes our state of knowledge in the future[1]. We may wish to take this effect on the forecasts for subsequent periods into account when planning which feeders to connect. In order to do this, we need to model how we expect our state of knowledge in the future, that is the accuracy of our forecasts, to be affected by our immediate decisions. One way to do this is to formulate the problem in a way that includes the subsequent periods and make the forecasts for the later periods in the model endogenously determined by the decisions which must be implemented immediately. The observations resulting from these decisions are uncertain at this

1. It is not necessary to claim that the decision variables affect the underlying stochastic process that defines feeder demand. Rather, we only need to claim that they affect our state of knowledge of the demand by changing which values we observe.

stage, so the future forecasts are stochastic. We now consider how to formulate this problem as a multi-stage stochastic program.

Let Θ_t represent all the information available by the end of period t. We use this information to construct the vector of means and variances of for each feeder at each time in the next decision interval. These vectors are now functions of Θ_t.

Define: $SKP(\boldsymbol{x}_t, \Theta_{t-1}) = (\boldsymbol{r} \circ \boldsymbol{\mu}_t(\Theta_{t-1}))^T \boldsymbol{x}_t - p\mathbb{E}[(D(\boldsymbol{x}|\Theta_{t-1}) - C)^+]$

If we ignore the future then at time t our problem is to $\max_{\boldsymbol{x}_t} SKP(\boldsymbol{x}_t, \Theta_{t-1})$. Θ_{t-1} is known at time t so the mean and variance of each feeder, $\boldsymbol{\mu}_t(\Theta_{t-1})$ and $\boldsymbol{\sigma}_t^2(\Theta_{t-1})$ are also known. We could solve this as a stochastic knapsack problem. We call this method of deriving a policy, and the policy obtained from it *myopic*. When the feeder scheduling problem is formulated considering only a single period the demand forecasts that form the data of our problem are exogenous to the model.

Now we consider the effect of our decisions at time t on the problem in future periods. The problem faced at time $t+1$ depends on Θ_t. Let $Pr(\Theta_t|\boldsymbol{x}_t, \Theta_{t-1})$ be the probability density function of Θ_t. Since the probability density of Θ_t is conditional upon $\boldsymbol{x}_t$, the decisions $\boldsymbol{x}_t$ influence the expected value of the knapsack problem at time $t+1$. The same argument can be made for all subsequent periods.

In general, it may be the case that the DSO has to 'fix' the decisions for several periods in advance using forecasts for multiple periods based on the same information. We can model this by partitioning the set of time periods $\mathcal{P}$ into a sequence of *decision intervals*. Let $\mathcal{T}$ denote a subset of $\mathcal{P}$ containing all the periods in a given decision interval.

Let $\boldsymbol{x}_\mathcal{T}$ refer to the set of decisions $\boldsymbol{x}_t$ for all time periods $t \in \mathcal{T}$ and $\Theta_\mathcal{T}$ be the information available at the end of the last time period in $\mathcal{T}$

After each decision interval the demand on each feeder can be observed in the periods it was connected. This information is incorporated into the forecasts for future periods. However, the forecasts for demand at all times *within* a particular decision interval are made based on the same set of information that was available at the beginning of the decision interval. Therefore the decisions in decision interval $\mathcal{T}$ do not affect the variance in $\mathcal{T}$ so the expectation of the sum of the penalties is the sum of the expected penalties in each period $t \in \mathcal{T}$[2].

The infinite horizon Feeder Scheduling Problem can be defined by the following recursion:

$$FSP(\Theta_{\mathcal{T}-1}) = \max_{\boldsymbol{x}_\mathcal{T}} SKP(\boldsymbol{x}_\mathcal{T}, \Theta_{\mathcal{T}-1}) + \int Pr(\Theta_\mathcal{T}|\boldsymbol{x}_\mathcal{T}, \Theta_{\mathcal{T}-1}) FSP(\Theta_\mathcal{T}) \mathrm{d}\Theta_\mathcal{T} \tag{4.4}$$

where:

2. This claim still holds if the errors within decision intervals are correlated. We elaborate on this claim in appendix 11.

$$SKP(\boldsymbol{x}_{\mathcal{T}}, \Theta_{\mathcal{T}-1}) = \sum_{t \in \mathcal{T}} (\boldsymbol{r} \circ \boldsymbol{\mu}_t(\Theta_{\mathcal{T}-1}))^T \boldsymbol{x}_t - pL(\frac{C - \boldsymbol{\mu}_t(\Theta_{\mathcal{T}-1})^T \boldsymbol{x}_t}{\sqrt{\boldsymbol{\sigma}_t^2(\Theta_{\mathcal{T}-1})^T \boldsymbol{x}_t^2}}) \sqrt{\boldsymbol{\sigma}_t^2(\Theta_{\mathcal{T}-1})^T \boldsymbol{x}_t^2} \quad (4.5)$$

To make this infinite recursion tractable we can terminate the recursion after some finite horizon h. We call this the *Finite Horizon Feeder Scheduling Problem*. We can find a (possibly sub-optimal) solution to the infinite horizon Feeder Scheduling Problem by solving the finite horizon Feeder Scheduling Problem for a finite set of decision intervals, implementing the decisions for the first decision interval and then re-solving for the next set of decision intervals.

4.3.1 Using State Space Models in the Feeder Scheduling Problem

Above we defined the Feeder Scheduling Problem with the mean and variance of total demand given as functions of the knapsack decision variables. In order to give a concrete specification of this problem we need to specify the model that produces forecasts of demand based on the observations. We now show how this can be done assuming that each feeder evolves according to any LGSSM. An example would be if the forecasts come from the ASM method developed in Chapter 3.

It will be helpful to recall the general linear Gaussian state space model:

$y_t = Z_t \alpha_t + \varepsilon_t$ (*observation* equation)

$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t$ (*state transition* equation)

with $\varepsilon_t \sim N(0, H_t)$ and $\eta \sim N(0, Q_t)$

Z_t is called the *design matrix*, T_t is called the *transition matrix* and R_t is called the *selection matrix*. y_t is the *observation* while $Z_t \alpha_t$ is referred to as the *signal* of the process.

For a known initial state (or known probability distribution of the initial state) and known variances for the state disturbance terms and observation error we can compute the maximum likelihood estimates for the state at time t conditional on all the prior data, $a_t = E[\alpha_t | y_1, ... y_{t-1}]$ using the Kalman filter. The signal of the process can be estimated by $Z_t a_t$.
The Kalman Filtering recursion is given by the equations:

Forecast error: $v_t = y_t - Z_t a_t$

Forecast error covariance matrix: $F_t = Z_t P_t Z_t' + H_t$

Kalman gain matrix: $K_t = T_t P_t Z_t' F_t^{-1}$

Kalman update: $L_t = T_t - K_t Z_t$ ($L_t = T_t$ if forecasting or observation missing)

Estimated state: $a_{t+1} = T_t a_t + K_t v_t$ ($a_{t+1} = T_t a_t$ if forecasting or observation missing)

State estimation error covariance matrix: $P_{t+1} = T_t P_t L_t' + R_t Q_t R_t'$

Note that a_t, P_t and F_t are calculated based on all observations available before time t. We will use the notation $X_{t|s}$ to mean the estimate of X at time t given information up to and including time s. We adopt the convention that $X_t = X_{t|t-1}$.

Forecasting with the Kalman Filter

Suppose we have a time series which we have observed up to time s and we want to forecast up to some forecast horizon h. That is, we want forecasts for all times $t : s < t < s+h$. Let's assume we have used the Kalman filter to compute the estimated state a_s and the state estimation error covariance matrix P_s.

We can forecast the time series by computing the expected value of the state, α_t, and multiplying by Z_t. To do this we project the estimated state a_t forwards recursively by multiplying the estimated state by the transition matrix T_t to get the expected value of the state at time $t+1$

Forecast Uncertainty

The error of our forecast will be given by the error in our estimation of the state and the observation noise. The contribution to the forecast error from the state estimation error is given by $Z_t P_t Z_t'$. This is independent of the observation noise which has known variance denoted H_t. The Forecast error covariance matrix is therefore given by $F_t = Z_t P_t Z_t' + H_t$.

Recall that if an observation is made at time t then for known P_t we update as follows:

$P_{t+1} = T_t P_t L_t' + R_t Q_t R_t'$

where: $L_t = T_t - K_t Z_t$,

$K_t = T_t P_t Z_t' F_t^{-1}$,

and: $F_t = Z_t P_t Z_t' + H_t$

If no observation is made at t then instead we update according to:

$P_{t+1} = T_t P_t T_t' + R_t Q_t R_t'$

Suppose $x_t = 1$ if an observation was made at t and 0 otherwise. We can then formulate the updated state estimation covariance matrix as:

$P_{t+1} = T_t P_t (T_t - x_t K_t Z_t)' + R_t Q_t R_t'$

We make the following observation:

P_{t+1} depends on P_t and whether or not an observation was made at time t, but it does not depend on what the value of the observation was. P_t is not a random variable. Therefore we can calculate $P_t = P_{t|s}$ using P_s for any $t > s$ as long as we know which observations will be made between s and t. It is not necessary to know the value of these observations.

The Feeder Scheduling Problem Using State Space Models for Demand Forecasts

In the state space case the information needed to evaluate $\boldsymbol{\mu}_t(\Theta_{\mathcal{T}-1})$ is the estimated state of every feeder at time s where s is the last period in decision interval $\mathcal{T}-1$. We then use the method for forecasting described in Section 4.3.1 to predict the state at time t and calculate the mean conditional on that state. Specifically, for each feeder $\mu_t = Za_t$ where Z is given by the state space model that defines the feeder's demand process.

To evaluate $\sigma_t^2(\Theta_{\mathcal{T}-1})$ we need to calculate the state estimation covariance matrix P_t for each feeder so that we can calculate $Z_tP_tZ_t^T + H_t$.

Let's denote by $a_{\mathcal{T}}$, $P_{\mathcal{T}}$ the sets containing a_t and P_t for all feeders and all time periods in $\mathcal{T}$ calculated using the information in $\Theta_{\mathcal{T}-1}$. These contain the necessary information to calculate the column vectors formed by stacking for all feeders the mean demand and forecast error variance respectively, which we will denote by $\boldsymbol{\mu}_t(\Theta_{\mathcal{T}-1}) = [Za_t]$ and $\boldsymbol{\sigma}_t^2(\Theta_{\mathcal{T}-1}) = [Z_tP_tZ_t^T + H_t]$

In section 4.3.1 we observed that the state estimation error covariance matrix is not a random variable and can be updated exactly given that we know which feeders we will connect (i.e. at which points we will observe the series). The integral in FSP can therefore be rewritten just as an integral over the future state. We adopt the notation $P_{\mathcal{T}+1}(P_{\mathcal{T}}, \boldsymbol{x}_{\mathcal{T}})$ to denote the state estimation error covariance matrices in decisions interval $\mathcal{T}+1$ calculated by updating $P_{\mathcal{T}}$ based on the decision taken in $\mathcal{T}$, $\boldsymbol{x}_{\mathcal{T}}$. Therefore, we can formulate the objective of the Feeder Scheduling Problem as:

$$FSP(a_{\mathcal{T}}, P_{\mathcal{T}}) = \max_{\boldsymbol{x}_{\mathcal{T}}} SKP(a_{\mathcal{T}}, P_{\mathcal{T}}, \boldsymbol{x}_{\mathcal{T}})$$

$$+ \int Pr(a_{\mathcal{T}+1}|a_{\mathcal{T}}, P_{\mathcal{T}}, \boldsymbol{x}_{\mathcal{T}}) FSP(a_{\mathcal{T}+1}, P_{\mathcal{T}+1}(P_{\mathcal{T}}, \boldsymbol{x}_{\mathcal{T}})) \mathrm{d}a_{\mathcal{T}+1}$$

$$SKP(a_{\mathcal{T}}, P_{\mathcal{T}}, \boldsymbol{x}_{\mathcal{T}}) = \sum_{t \in \mathcal{T}} (R \circ [Za_t])^T \boldsymbol{x}_t - pL\left(\frac{C - [Za_t]^T \boldsymbol{x}_t}{\sqrt{[Z_tP_tZ_t^T + H_t]^T \boldsymbol{x}_t^2}}\right) \sqrt{[Z_tP_tZ_t^T + H_t]^T \boldsymbol{x}_t^2}$$

The Feeder Scheduling Problem is a multi-stage stochastic program with integer decision variables and a high dimensional state space. In general, this problem class is very difficult to solve. In particular solving $FSP(a_{\mathcal{T}}, P_{\mathcal{T}})$ is problematic for 2 reasons

1. Evaluating the future state estimation covariance matrix function $P_{\mathcal{T}+1}(P_{\mathcal{T}}, \boldsymbol{x}_{\mathcal{T}})$
2. The need to integrate over the future state $a_{\mathcal{T}+1}$

We now discuss the case in which the feeder demand is modelled by an AR(1) process around a deterministically evolving mean. This simplifies evaluating the future state estimation covariance matrix substantially and is a good approximation to the demand models we developed in Chapter 3.

We will then consider whether we can get an approximate solution to FSP without the need to evaluate the integral. We discuss some bounds on the true solution and evaluate these empirically.

Feeder Scheduling Problem with Autoregressive Model for Demand

In the general case it is difficult to include the calculation of P_t in the feeder scheduling problem because it involves the inversion of the matrix F_t (to calculate the Kalman gain matrix K_t) and a series of potentially large matrix multiplications.

However, many time series of practical interest can be modelled by an AR(1) process around a deterministically evolving mean. When we fit the ASM models to real feeder data, the variance of the state disturbance term corresponding to the AR(1) term is orders of magnitude larger than the other state disturbances. The observation error is also small relative to the autoregressive error. This means that over the short term the demand on each feeder can be approximated by an AR(1) plus white noise process around a deterministically evolving mean. In this case the calculation of P_t can be simplified substantially.

In state space form this model is given by:

$y_t = \alpha_t + \varepsilon_t$

$\alpha_{t+1} = \phi x_t + \eta_t$

$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, $\eta_t \sim N(0, \sigma_\eta^2)$

Evaluating the future state estimation covariance matrix function $P_{\mathcal{T}+1}(P_{\mathcal{T}}, \boldsymbol{x}_{\mathcal{T}})$ is substantially simplified by this model as we show below.

The state space matrices defining the model are $Z = 1, T = \phi, H = \sigma_\varepsilon^2, Q = \sigma_\eta^2, R = 1$. This process is stationary as long as $\phi \in [-1, 1]$.

If we apply the Kalman filter to this model the forecast uncertainty is given by $F_t = Z_t P_t Z_t' + H_t = P_t + \sigma_\varepsilon^2$ which requires us to update P_{t+1} recursively. Suppose P_t is given. Using the Kalman filtering equations we have:

$P_{t+1} = \phi P_t(\phi - x_t K_t) + \sigma_\eta^2$

where x_t is a binary variable taking the value 1 if an observation is made at time t and 0 otherwise.

Therefore if no observation is made at time t, $P_{t+1} = \phi^2 P_t + \sigma_\eta^2$. By recursive substitution we can show that if the process has not been observed for n periods P_{t+n} is given by:

$P_{t+n} = \phi^{2n} P_t + \sum_{i=1}^{n} \phi^{2(i-1)} \sigma_\eta^2$

Now let's consider how we update P_{t+1} if an observation is made at time t. For this model $F_t = P_t + \sigma_\varepsilon^2$, therefore $K_t = \frac{\phi P_t}{P_t+\sigma_\varepsilon^2}$. Therefore, if an observation is made at time t then:

$P_{t+1} = \phi P_t(\phi - \frac{\phi P_t}{P_t+\sigma_\varepsilon^2}) + \sigma_\eta^2$

If σ_ε^2 is small relative to P_t then $K_t \approx \phi$ and $P_{t+1} \approx \sigma_\eta^2$. This is a reasonable assumption based on our experience fitting structural models to actual demand data. Based on this we propose the following simplified model of the variance:

If $x_t = 1$, $P_{t+1} = \sigma_\eta^2$

If $x_t = 0$ for n consecutive periods after time $t-1$,

$P_{t+n} = \sum_{i=0}^{n} \phi^{2i}\sigma_\eta^2$

In other words the forecast error variance grows according to the sum of a geometric series and reverts to the variance of the white noise error term after an observation is made.

4.3.2 Approximate Solutions to Feeder Scheduling Problem

By solving $FSP(a_\mathcal{T}, P_\mathcal{T})$ we obtain the optimal policy for the first decision interval which we denote $\hat{\boldsymbol{x}}_\mathcal{T}$. The expected value of implementing these decisions is $SKP(a_\mathcal{T}, P_\mathcal{T}, \hat{\boldsymbol{x}}_\mathcal{T})$. The actual value of these decisions is only known when the stochastic parameters are realized and is the sum of the realized rewards and penalties, i.e. $\sum_{t\in\mathcal{T}} R_t(\hat{\boldsymbol{x}}_t) - p(D_t(\hat{\boldsymbol{x}}_t) - C)^+$.

Over the long term the average expected objective should equal the average "predicted" objective, $SKP(a_\mathcal{T}, P_\mathcal{T}, \hat{\boldsymbol{x}}_\mathcal{T})$.

Although the solution to $FSP(a_\mathcal{T}, P_\mathcal{T})$ is the optimal policy, it is very difficult computationally. However, there are many other ways of obtaining a (possibly sub-optimal) policy. We can make the same comparison between the expected objective according to the chosen solution method to the problem and the actual realized objective. For any solution method that does not exactly solve $FSP(a_\mathcal{T}, P_\mathcal{T})$ the long term average of the actual realized objective will not equal the expected value. Figure 4.1 represents schematically the difference between the predicted and realized objective value for various approaches to the problem. We now discuss these approaches and argue that we can use this to bound the possible window in which the average value of exactly solving FSP can lie.

One approach to obtaining a policy is the Perfect Foresight policy, denoted PF in Figure 4.1. The Perfect Foresight policy is obtained by solving a deterministic knapsack policy in every stage with the true realized demand on each feeder. Since there is no difference between the demand in the optimization problem and the realized demand the predicted and realized objectives are identical. No feasible policy can improve on the Perfect Foresight policy.

Alternatively, we could solve a deterministic knapsack problem on the assumption that the knapsack items have a known size given by the forecasts at this stage. For each decision interval

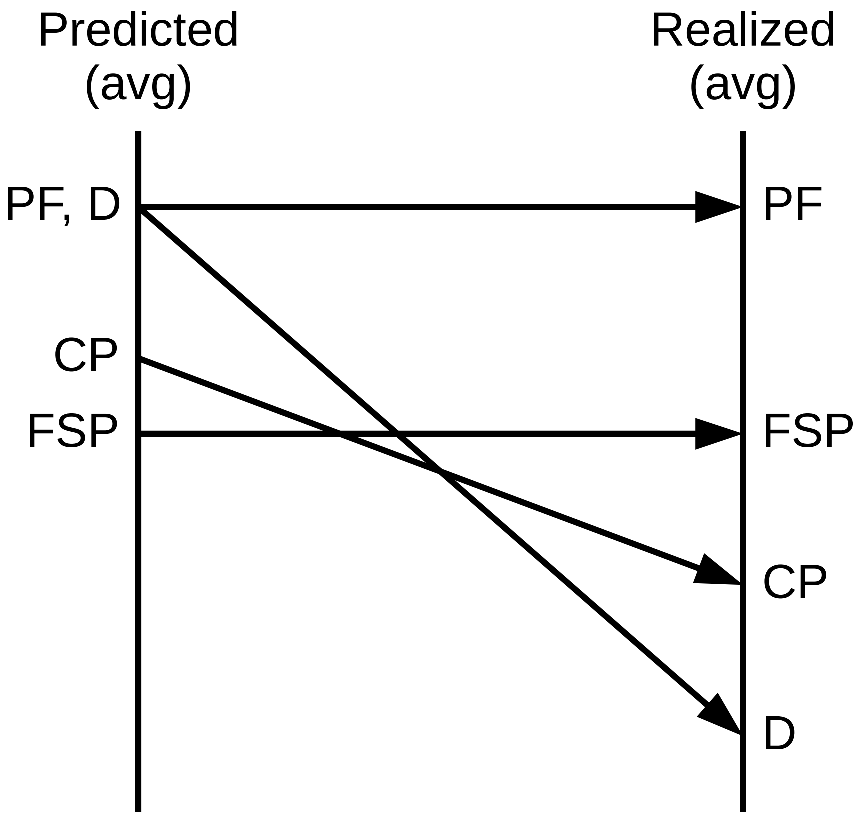


Figure 4.1: A comparison of the relative “predicted” and “true” objective values from various methods of deriving a policy for the feeder scheduling problem: PF - perfect foresight, D - deterministic, CP - central path, FSP - feeder scheduling problem

we implement the first stage decisions of the solution to a multi-stage deterministic problem[3]. On average we would expect the predicted objective function value of the deterministic problem to be similar to the Perfect Foresight problem. However, the realized objective function of the deterministic solution should be worse than the predicted objective.

Furthermore, on average the realized objective of the deterministic policy will be worse than FSP because the assumption of deterministic demand means that the deterministic policy will tend to connect too many feeders given the actual uncertainty about demand. The long-term average objective value of FSP should therefore fall somewhere between the long term average realized objective of the deterministic policy and the long term average predicted/realized objective of the Perfect Foresight solution.

We may be able to obtain a tighter bound as follows.

Recall that $a_{\mathcal{T}}$ is the sets containing the estimated state a_t for all feeders and all time periods in $\mathcal{T}$ calculated using the information available at the end of the previous decision interval, i.e. $\Theta_{\mathcal{T}-1}$. We adopt the notation $a_{\mathcal{T}|\mathcal{T}'}$ to denote the sets containing the estimated state a_t for all feeders and all time periods in $\mathcal{T}$ calculated using the information available at the end of a particular decisions interval, $\mathcal{T}'$.

We can approximate FSP by CP, the *Central Path* of the feeder scheduling problem:

$$FSP(a_{\mathcal{T}'}, P_{\mathcal{T}'}) \approx CP(a_{\mathcal{T}'|\mathcal{T}'-1}, P_{\mathcal{T}'}) = \max_{\boldsymbol{x}_{\mathcal{T}'}} SKP(a_{\mathcal{T}'}, P_{\mathcal{T}'}, \boldsymbol{x}_{\mathcal{T}'}) + CP(a_{\mathcal{T}+1|\mathcal{T}'-1}, P_{\mathcal{T}'+1}(P_{\mathcal{T}'}, \boldsymbol{x}_{\mathcal{T}'}))$$

where:

$$CP(a_{\mathcal{T}|\mathcal{T}'}, P_{\mathcal{T}}) = \max_{\boldsymbol{x}_{\mathcal{T}}} SKP(a_{\mathcal{T}|\mathcal{T}'}, P_{\mathcal{T}}, \boldsymbol{x}_{\mathcal{T}}) + CP(a_{\mathcal{T}+1|\mathcal{T}'}, P_{\mathcal{T}+1}(P_{\mathcal{T}}, \boldsymbol{x}_{\mathcal{T}}))$$

In CP, compared to FSP, we do not integrate over possible future values of $a_{\mathcal{T}}$, rather we assume that $a_{\mathcal{T}}$ will evolve according to our best forecast using the information available at the beginning of $\mathcal{T}'$.

The long-term average predicted objective from the Central Path problem will probably be higher than FSP because it does not "hedge" against uncertainty about the future predicted state. For the same reason the long-term average realized objective from the Central Path problem is expected to be lower than the FSP. In principle it is possible that mean value of the $a_{\mathcal{T}}$ will be worse suited for the packing problem in SKP than the values close to the mean and so the

3. Note that for the deterministic problem the only way that the problems of each stage could depend on one another is if there are supply quality constraints. Without supply quality constraints we could just solve a single stage problem.

expected value of CP will be worse than FSP. However, we think that over the long term this effect will be minimal.

If accounting for uncertainty about the future predicted state is not important then we would expect the gap between the long-term average predicted and realized objective of CP to be small. This depends on the specific facts about the data and distribution of items. We will empirically investigate this by carrying out simulations that reflect the likely range of parameter for the feeder scheduling problem faced by DSOs.

A Note on Stochastic Supply

One area we have not explored in detail is uncertainty about the level of supply available from the grid. As observed in Section 4.2 this can be incorporated in to the model easily by including an additional item in the problem with mean weight 0, variance equal to the variance of the supply and the decision variable for this item fixed so that it must be selected at every period. It is therefore easier to model the future forecast variance of the supply compared to demand because this is always observed and so is not dependent on the decisions. A more complex state space model for the capacity level could be adopted rather than the deterministic mean plus AR(1) model proposed for the individual feeders. If the uncertainty in the supply is large relative to the uncertainty about each feeder's demand, then the future problem may be very different depending on how the supply variable evolves. Consequently, it may be more important to integrate over the future state of the problem. However, because this uncertainty is only with respect to a single dimension, the supply level, then this may be a tractable problem to solve with a scenario tree for the supply level.

4.4 Model Formulation

We can now formulate the Central Path approximation to FSP as a Mixed Integer Linear Program using the sets and parameters in Table 4.2

Where a set is ordered we use the notation $\inf\{\cdot\}$ and $\sup\{\cdot\}$ to denote the first and last element of the set respectively.

We use the notation $X \setminus Y$ to denote the set containing all the elements of X except any that are elements of Y

Table 4.2: Sets, parameters and variables used to define the feeder scheduling problem

	Notes	Description
Sets		
$\mathcal{F}$		Feeders
$\mathcal{P}$	ordered	Time periods
$\mathcal{I}$	ordered	Decision intervals
$\mathcal{M}$		Minimum connection frequency intervals
$\mathcal{T}_i \subset \mathcal{P} \forall j \in \mathcal{I} \cup \mathcal{M}$	ordered	Periods in decision interval i
$\mathcal{S}_t := \{t' \in \mathcal{P} : t' < t \leq t' + m\}$		Time slices for max outage constraint
$\mathcal{S}_{t \in \mathcal{T}} = \mathcal{S}_t \cap \mathcal{T}$		
$\mathcal{S}_{t < \mathcal{T}} = \mathcal{S}_t \setminus \mathcal{S}_{t \in \mathcal{T}}$		
$\mathcal{Y}$	ordered	set of indices for PWL approximation of mean
$\mathcal{Z}$	ordered	set of indices for PWL approximation of standard deviation
Parameters	**Indexing set**	
C		Demand theshold for overconsumption penalty
p		penalty for overconsumption
r_i	$\mathcal{F}$	reward per unit demand on feeder i
$\hat{\sigma}^2_{i,t}$	$\mathcal{F}$,$\mathcal{P}$	Forecast variance of feeder i at t
$\mu_{i,t}$	$\mathcal{F}$,$\mathcal{P}$	mean of feeder i at time t
m_y	$\mathcal{Y}$	Points in approximation of mean
$L(m_y)$	$\mathcal{Y}$	Points in approximation of expected penalties
v_z	$\mathcal{Z}$	Points in approximation of σ^2
s_z	$\mathcal{Z}$	Points in approximation of σ
ϕ_i	$\mathcal{F}$	Autoregression coefficient

	Notes	Description
σ_i^2	$\mathcal{F}$	Variance of ar disturbance for feeder i
N_i^0	$\mathcal{F}$	Number periods feeder i was off prior to start
f		Minimum number of connections in $\mathcal{P}$
$\hat{f}_i = f - \sum_{t \in \mathcal{P} < \mathcal{T}} x_{i,t} - \sum_{t \in \mathcal{P} > \mathcal{T}} 1$	$\mathcal{F}$	Minimum number of connections for feeder i
m		Maximum outage length
$\delta_{i,t} = \sum_{t' \in \mathcal{S}_{t<\mathcal{T}}} x_{i,t'}$	$\mathcal{F}$	Max outage constraint indicator
Variables	**Indexing set**	
$x_{i,t} \in \{0,1\}$	$\mathcal{F}$,$\mathcal{P}$	Binary, 1 if feeder i connected at t, 0 otherwise
$\sigma_t^2(\boldsymbol{x}_t)$	$\mathcal{P}$	Connected demand variance at time t
$\mu_t(\boldsymbol{x}_t)$	$\mathcal{P}$	Connected demand variance at time t
$r_t(\boldsymbol{x}_t)$	$\mathcal{P}$	Expected revenue at time t
$\sigma_{i,t}^2$	$\mathcal{F}$,$\mathcal{P}$	Contribution to total variance from feeder i at t
$N_{i,t}$	$\mathcal{F}$,$\mathcal{P}$	Periods feeder i is disconnected prior to t
$\nu_{i,t}$	$\mathcal{F}$,$\mathcal{P}$	Periods diconnected prior to decision interval plus time in decision interval
$\delta_{i,t,k} \in \{0,1\}$	$\mathcal{F}$,$\mathcal{P}$	indicator variable denoting if feeder i is disconnected for k periods before t
$\lambda_{y,t} \in [0,1]$	$\mathcal{Y}$,$\mathcal{P}$	weight variable for approximation of mean penalties
$\pi_{z,t} \in [0,1]$	$\mathcal{Z}$,$\mathcal{P}$	weight variable for approximation of standard deviation
$\alpha_{z,t} \in \{0,1\}$	$\mathcal{Z} \setminus \inf\{\mathcal{Z}\}$,$\mathcal{P}$	indicator variable for SOS2 sets

Equation 4.6 defines the objective function to be maximised, i.e. rewards minus expected penalties. 4.7 to 4.9 define a piecewise linear inner approximation to the expected penalties as discussed in 4.2.2.

4.10 defines the knapsack variance as the sum of the variance contributed by each feeder. The variance contributed by each feeder is defined according to 4.11 in the first decision interval, i.e. it is given by the forecast variance $\hat{\sigma}^2_{i,t}$ if the feeder is connected, 0 other wise. In subsequent decision intervals the contribtion to the knapsack variance calculated according to how long the feeder has been disconnected in 4.12, but is 0 if the feeder is disconncted in this period. To avoid multiplying the right hand side of the constraint by $x_{i,t}$ we use the "big M" method to linearise this constraint.

Constraints 4.13 to 4.18 measure the length of time that each feeder is disconnected. $\delta_{i,t,k}$ is 1 if the feeder was disconnected for at least k periods prior to time t, 0 otherwise. $N_{i,t}$ measures how many periods feeder i was disconnected consecutively prior to t. $\nu_{i,t}$ measures how many periods feeder i was disconnected for prior to the start of the decision interval containing t plus the number of time periods since the start of that decision interval. This is needed because we only have access to obsevations *before* the start of the decision interval to update the forecasts.

4.19 to 4.21 define a piecewise linear approximation of the standard deviation of the total demand in terms of the variance of the total demand, which is the sum of the variance of the selected feeders. 4.22 to 4.25 define SOS2 sets to ensure that only two adjacent weight variables $\pi_{z,t}$ can be non zero.

In Section 4.4.1 we define a minimum connection frequency constriants using 4.26. In Section 4.4.2 we define maximum outage length constraints using constraint 4.29.

Cost Function

$$\max \sum_{t\in\mathcal{P}} \left(\sum_{i\in\mathcal{F}}^{n} x_{i,t} r_i \mu_{i,t} - p \sum_{y\in\mathcal{Y}} \lambda_{y,t} L(m_y) \right) \tag{4.6}$$

Piecewise linear approximation of expected penalties

$$\lambda_{y,t} \geq 0 \tag{4.7}$$

$$\sum_{y\in\mathcal{Y}} \lambda_{i,t} = \sigma_t, \tag{4.8}$$

$$C - \sum_{i=0}^{n} x_{i,t} \mu_{i,t} = \sum \lambda_i m_i, \tag{4.9}$$

Define Feeder Variance

$$\sigma_t^2 = \sum_{i=0}^{n} \sigma_{i,t}^2, \tag{4.10}$$

$$\sigma_{i,t}^2 \geq \hat{\sigma}_{i,t}^2 x_{i,t}, t \in \mathcal{T}_1, \tag{4.11}$$

$$\sigma_{i,t}^2 \geq \sum_{k=0}^{\kappa} \delta_{i,t,k} \phi^{2k} \sigma_\eta^2 - M(1 - x_{i,t}), t : \exists j \in \mathcal{I} : t \in \mathcal{T}_j \tag{4.12}$$

Measure disconnection length

$$N_{i,t} \geq N_{i,t-1} + 1 - x_{i,t-1} M, \forall t \in \mathcal{P} : t > 1, \tag{4.13}$$

$$N_{i,t} \geq N_{i,t}^0, t = 1, \tag{4.14}$$

$$\nu_{i,t} \geq N_{i,t}, \forall i \in \mathcal{F}, \forall t : \exists j \in \mathcal{I} : t \in \inf\{\mathcal{T}_j\}, \tag{4.15}$$

$$\nu_{i,t} \geq \nu_{i,t-1} + 1, \forall i \in \mathcal{F}, \forall t : \exists j \in \mathcal{I} : t \in \mathcal{T}_j \setminus \inf\{\mathcal{T}_j\}, \tag{4.16}$$

$$\nu_{i,t} \leq k - 1 + \delta_{i,t,k} M, \tag{4.17}$$

$$\sum_{t' \in \mathcal{S}_t} x_{i,t'} + \delta_{i,t} \geq 1 \forall t \in \mathcal{P}, i \in \mathcal{F} \tag{4.18}$$

Piecewise linear approximation of standard deviation

$$\sigma_t^2 \leq \sum_{z \in \mathcal{Z}} \pi_{z,t} v_z, \tag{4.19}$$

$$\sigma_t \geq \sum_{z \in \mathcal{Z}} \pi_{z,t} s_z, \tag{4.20}$$

$$\sum_{z \in \mathcal{Z}} \pi_{z,t} = 1, \tag{4.21}$$

$$\sum_{z \in \mathcal{Z} \setminus \inf\{\mathcal{Z}\}} \alpha_{z,t} = 1, \tag{4.22}$$

$$\sum_{y \in \mathcal{Y}} \pi_{y,0,t} \leq \alpha_{1,t}, \tag{4.23}$$

$$\sum_{i} \pi_{y,z,t} \leq \alpha_{z,t} + \alpha_{z+1,t}, \tag{4.24}$$

$$\sum_{i} \pi_{i,\sup\{\mathcal{Z}\},t} \leq \alpha_{\sup\{\mathcal{Z}\},t}, \tag{4.25}$$

4.4.1 Minimum Connection Frequency Constraint

We might be required to satisfy a minimum connection frequency for feeder i which can be expressed by the inequality:

$f \leq \sum_{t \in \mathcal{P}} x_{i,t}$

Suppose that we are optimising over decisions that take place in $\mathcal{T}$, which is a subset of $\mathcal{P}$. We want to replace this constraint with a modified constraint that only includes $\boldsymbol{x}_{\mathcal{T}}$ but ensures that for all $\boldsymbol{x}_{\mathcal{T}}$ there exists some possible future set of decisions such that the constraint is satisfied.

Let $\boldsymbol{x}_{\mathcal{P}}$ refer to the decisions in all periods in $\mathcal{P}$. Similarly let $\boldsymbol{x}_{\mathcal{T}}$ refer to the decisions in all periods in $\mathcal{T}$. Denote by $\boldsymbol{x}_{\mathcal{P}<\mathcal{T}}$ and $\boldsymbol{x}_{\mathcal{P}>\mathcal{T}}$ the decisions taken in all periods before and after $\mathcal{T}$ respectively.

We can subtract from f the number of times the feeder was connected before $\mathcal{T}$ and the maximum number of periods it could be connected after $\mathcal{T}$. The modified left hand side of the constraint is the minimum number of connections in $\mathcal{T}$ that the feeder must be connected:

$$\hat{f}_i = f - \sum_{t \in \mathcal{P}<\mathcal{T}} x_{i,t} - \sum_{t \in \mathcal{P}>\mathcal{T}} 1 \leq \sum_{t \in \mathcal{T}} x_{i,t} \forall i \in \mathcal{F}, \tag{4.26}$$

4.4.2 Maximum Outage Constraint

Alternatively suppose m is the maximum allowable outage length in $\mathcal{P}$.

The maximum outage length constraint can be expressed by exhaustively dividing P into "slices" of length $m+1$ and requiring that each feeder is connected at least once in each slice:

$$\sum_{t' \in \mathcal{S}_t} x_{i,t'} \geq 1, \tag{4.27}$$

$$\mathcal{S}_t := \{t' \in \mathcal{P} : t' \leq t \leq t' + m\} \forall t \in \mathcal{P} \tag{4.28}$$

Suppose $\mathcal{S}_t$ includes time periods after the current decision interval $\mathcal{T}$. In this case we can ignore the constraint because it will always be possible to satisfy it after the current decision interval. This is simply achieved by redefining the set of "slices" to only include slices $\mathcal{S}_t$ that end in the current decision interval:

$$\mathcal{S}_t := \{t' \in \mathcal{P} : t' \leq t \leq t' + m\} \forall t \in \mathcal{T}$$

For slices that include time periods that happened before $\mathcal{T}$ we can ignore the constraint if the feeder was connected in that slice before $\mathcal{T}$. To impose this condition for each feeder i and time period t we can define a parameter $\gamma_{i,t}$ which indicates if feeder i was connected before the start of $\mathcal{T}$ in the last m periods prior to t:

$\gamma_{i,t} = \sum_{t' \in \mathcal{S}_{t<\mathcal{T}}} x_{i,t'}$

Where $\mathcal{S}_{t<\mathcal{T}}$ is the set of periods in this slice $\mathcal{S}_t$ before the current decision interval $\mathcal{T}$

The maximum outage constraint can therefore be formulated:

$$\sum_{t' \in \mathcal{S}_t} x_{i,t'} + \gamma_{i,t} \geq 1 \forall t \in \mathcal{T}, i \in \mathcal{F} \tag{4.29}$$

4.5 Software implementation

We formulate the model from Section 4.4 in AMPL. The resulting MILP problem is solved using CPLEX. In the AMPL formulation we explicitly declared an SOS2 set to ensure that only two adjacent weight variables $\pi_{z,t}$ can be non zero rather than using the constraints 4.22 to 4.25. This allows CPLEX to exploit the special structure of the problem using specialized branching strategies for SOS2 sets. The problem was solved using a Dell PowerEdge R920 with Four Intel Xeon E7-4830 v2 2.2GHz, 20M Cache, 7.2 GT/s QPI, Turbo (4x10Cores) and 256 GB RAM.

4.6 Evaluation

In this section we perform some numerical experiments using the stochastic knapsack model to schedule feeders on a rolling horizon basis. We consider how a myopic policy compares to solving the feeder scheduling problem taking the effect on future periods into account. Since we cannot solve FSP directly we consider the central path approximation described in Section 4.3.2. We argued in Section 4.3.2 that the predicted objective value of the central path approximation gives us a good estimate of the upper bound on how much better we could do by solving FSP. The true objective value can be compared to the myopic solution to see how well this solution method does in practice.

Suppose we have a model of the demand process on each feeder. Given a policy, that is, a set of feeders to be connected, we can query the model to return the observed demand on each feeder. These observations become the input to a model that produces a forecast mean and variance for each feeder up to the end of the next decision interval or further if required. We use these forecasts as inputs for a model to determine the policy for the next decision interval. We can then iterate the process, by revealing demand on feeders connected according to the policy and updating our forecasts. A schematic representation of the process is given in Figure 4.2.

In order to implement this evaluation procedure, we need to define an appropriate demand model, forecast method and policy optimization procedure.

In Chapter 3 we fitted structural models to 76 feeders from Jos Electricity Distribution Company. Of these we selected 65 which seem to satisfy our assumption that AR variance is much

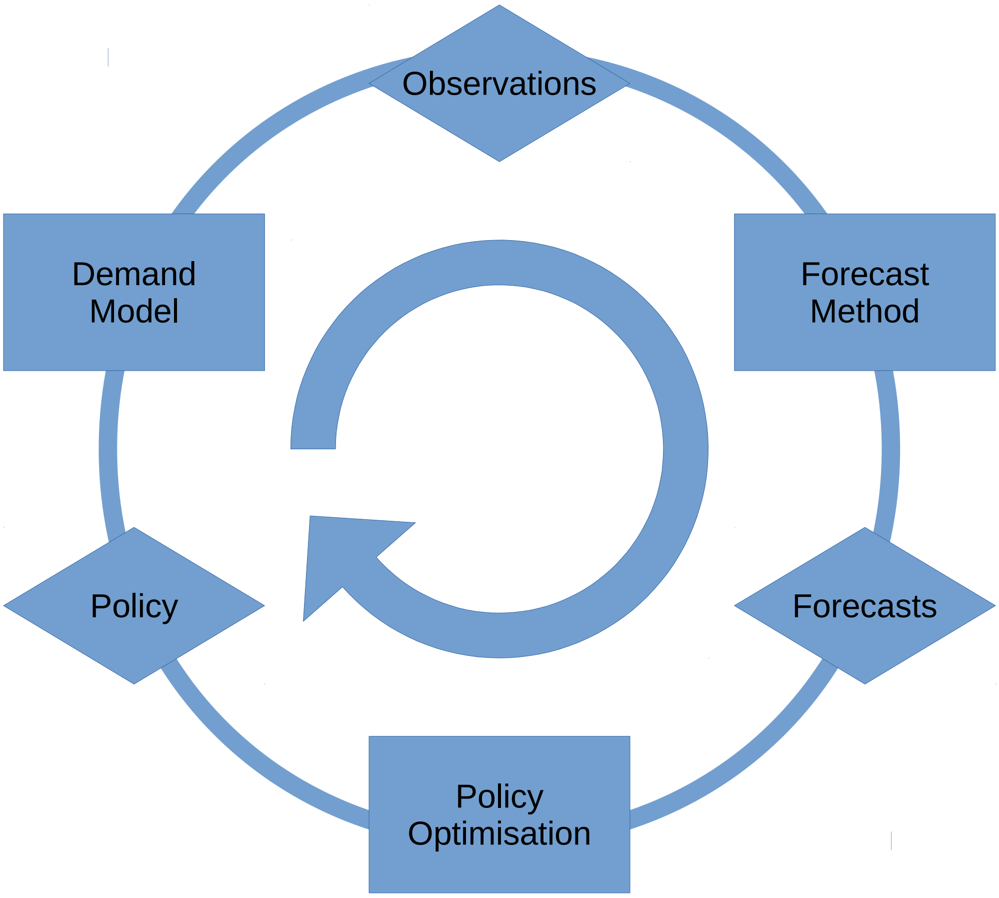


Figure 4.2: Combined Optimisation-Simulation model of the Feeder Scheduling Problem

larger than other variance terms[4]. We simulate the evolution of demand according to an AR(1) process using parameter values taken from the Jos models. We reduce the seasonal period from 24 hours to 6 hours. This allowed us to experiment with solving a problem where the optimization horizon extends over the whole seasonal period without having to solve a 24-period problem. We can in principle solve a problem of this length, but it would make it impractical to run the evaluation loop for enough iterations to get reliable results.

Since we have simulated the results from a known AR(1) process with a deterministic mean, we can use the optimal forecasts from the Kalman filter. We assume all feeders connected initially so we have starting forecasts of best possible quality.

Consider now the case where decisions are only needed for one period at a time and that after every period we can revise our decisions for the future – i.e. decisions intervals consist of one period. We vary the number of decision intervals that are included in the problem in addition to the first interval. We consider lengths for the rolling horizon of 1, 3 and 6 decision intervals ahead after the first stage decisions. Setting the horizon length to 0 corresponds to the Myopic problem. The penalty for overconsumption is 20 per unit. The maximum capacity is 4000 at all times. This corresponds to roughly 2/3 of off-peak demand and 2/5 peak demand which is a credible range for Nigeria.

We first consider a case without any supply quality constraints. We also consider how the reward per unit supply on each feeder affects our results by considering two cases. In case 1 the reward per unit supply is 1.5 for all feeders. In case 2 the reward per unit supply varies between feeders – Prior to solving the problem a reward level is determined for each feeder by sampling from a uniform distribution between 1 and 2. We then repeat these experiments with a constraint on the maximum outage length.

4.6.1 No Supply Quality Constraints

Table 4.3 shows the average realized objective value per period over a sample of 480 periods. For comparison we also show the predicted objective value.

The realized objective is not appreciably different from the predicted objective in either the equal reward per feeder or varying reward cases. According to our reasoning from Section 4.3.2 this suggests that we could not do any better by solving the true Feeder Scheduling Problem.

Neither the predicted or realized objective value is improved by increasing the length of horizon. This indicates that in this case there is no penalty to using a myopic policy. This is not true in general however; it depends on the characteristics of the stochastic demand. To illustrate

4. The other feeders are characterized by particularly large outliers or egregious structural breaks which suggest something unusual changed the pattern of demand over the period in question such as the connection or disconnection of groups of customers. These events would in practice be known about in advance and could be factored into forecasts.

	horizon	0	1	3	6
Rewards	obj				
equal	Predicted	5863.2	5863.5	5862.9	5864.6
	Realized	5862.5	5858.8	5854.0	5863.1
unequal	Predicted	6601.4	6599.4	6599.8	6598.2
	Realized	6615.0	6614.2	6614.0	6611.5

Table 4.3: Average predicted and realized objective value (i.e. expected rewards minus expected penalties vs actual rewards minus actual penalties) for Central Path approximation of feeder scheduling problem with no supply quality constraints

this, we next develop a family of problems which show that under some circumstances looking ahead to plan for the variance can result in a better policy than being myopic.

4.6.2 An Example Where Planning for the Variance Matters – AR(1) Rolling Horizon Model

Consider a case with n feeders, split into two groups, each consisting of half the feeders. Suppose that feeder demand follows an AR(1) process around a deterministically evolving and known mean. Assume that the variance of the noise process driving the evolution of each item is σ_A for the first group and σ_B for the second.

For this example, we consider a problem with just two periods. After the period 1 decisions are implemented it is possible to use the period 1 observations to revise the period 2 decisions. Recall that if a feeder is connected in period 1 then it's period two variance is just σ^2. However, if a feeder is not included then it's variance is $\sigma^2 + \phi^2\sigma^2$.

Suppose for group A feeders ϕ_A is close to 0 and for group B ϕ_B is close to 1. Therefore, if a feeder in group B is not connected in period 1 it's forecast error variance in period 2 will be almost twice that of a feeder in group A.

Denote the capacity threshold by C. Suppose that the deterministically evolving mean of each feeder is $z_1 = 2C/n$ in period 1, and $z_2 = C/n$ in period 2. In period 1 we have enough capacity to connect about half of the feeders. In period 2 the average demand is smaller, so we have enough capacity to connect almost all the feeders. The precise displacement of the optimal mean demand from the capacity depends on the relative size of the reward per unit demand and the penalty for overconsumption.

Suppose that the reward per unit supplied to group A is slightly greater than the reward for group B, e.g. $r_A = (1+\theta)r_B$. A myopic policy in this case is to connect (almost) all the feeders in group A in period 1 and to connect (almost) all the feeders of both groups in period 2.

However, a better policy would be to connect (almost) all the feeders in group B in period 1 and then connect (almost) all the feeders of both groups in period 2.

The disadvantage of the myopic policy is that the variance of the group B feeders grows at a faster rate than the group A feeders when disconnected. Therefore, in period 2, when we can connect most of the feeders, the total variance of the knapsack is higher if we follow the short-sighted policy than the alternative.

Simulations

Consider the case with the following parameters:

- n = 100
- C = 200
- $\sigma_A = \sigma_B = 1$
- $\phi_A = 0.01$
- $\phi_B = 0.99$
- $r_A = .2$
- $r_B = .2001$
- p = 1 (overconsumption penalty)

We simulate 100 cases of the evolution of the AR processes. In each case we find the 'lookahead' and myopic policies. We compare the expected rewards in each period, that is, the objective value according to the optimization problems to the actual objective value, that is the result of implementing the policy.

We define the Myopic and Lookahead policies as follows:

Myopic

- Implement a plan that maximizes expected rewards minus overconsumption penalties in period 1
- Observe value of connected feeder and update forecasts for period 2
- Implement a plan that maximizes expected rewards minus overconsumption penalties in period 2

look-ahead

- Implement the period 1 decisions of a plan that maximizes expected rewards minus overconsumption penalties in period 1 and 2 using the Central Path approximation to the feeder scheduling problem.
- Observe value of connected feeder and update forecasts for period 2
- Implement a plan that maximizes expected rewards minus overconsumption penalties in period 2

In Figure 4.3 we plot the average expected objective and "true" objective after each simulation to illustrate that our estimates of the expected objective and true objective in each case seem to have converged.

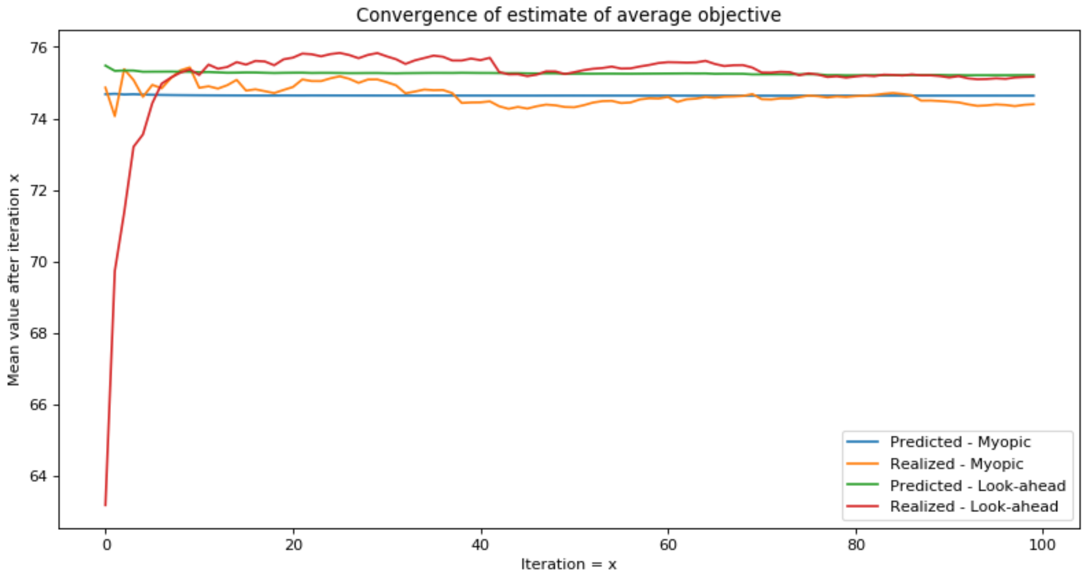


Figure 4.3: Convergence of average realized objective value to predicted objective function value

We can see that the lookahead policy is on average about 1.5% better than the myopic policy.

In practice the scheduling problem is repeated indefinitely. We simulate the evolution of the AR processes of 100 items as in the first case for 120 periods (5 days). We find the myopic policy for each hour separately as above but repeated for the whole duration. We also find a "look-ahead" policy as above, except that at every stage we are planning for two periods and implementing the first stage decisions.

In Figure 4.4 we plot the objective value over time of the myopic policy and lookahead policy. The oscillation of the predicted objective function is caused by the oscillation in the level of demand. When the demand is higher the number of feeders selected to use up the available capacity is lower. This means the knapsack variance is lower, so for the same expected penalty we can choose a knapsack mean closer to the capacity.

It is interesting to note that around the fourth day both policies perform worse than average but the lookahead policy is able to recover more quickly to the long-term average objective value. What seems to be happening is that around the fourth day demand was unusually high overall, so several items were disconnected for multiple periods. As a result, their forecast variance grew to the point that they became unattractive to connect at any time. The lookahead policy can account for the fact that by connecting these items in period 1 they will have a much lower variance in period 2. The myopic policy does not realize this and suffers a continual penalty to the objective after this point.

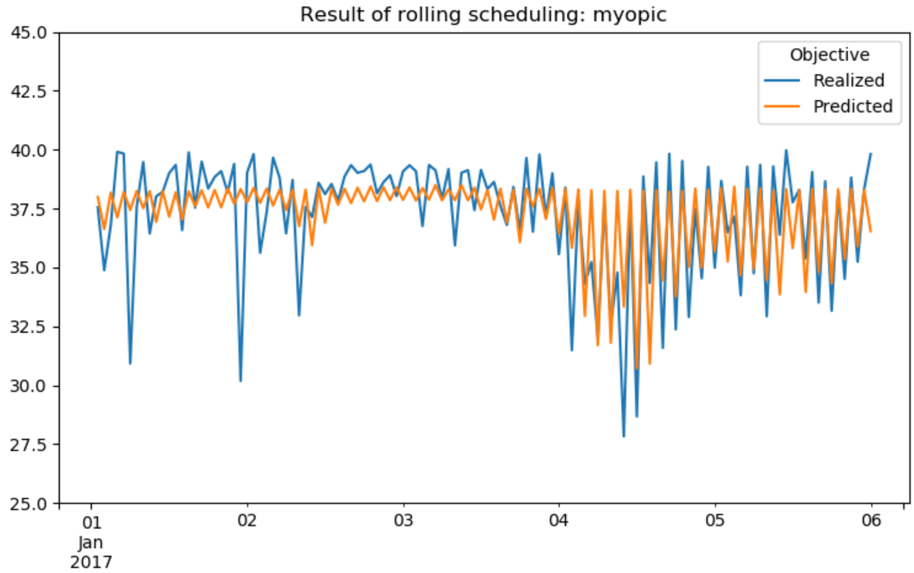


Figure 4.4: Predicted and realized objective function value from myopic scheduling

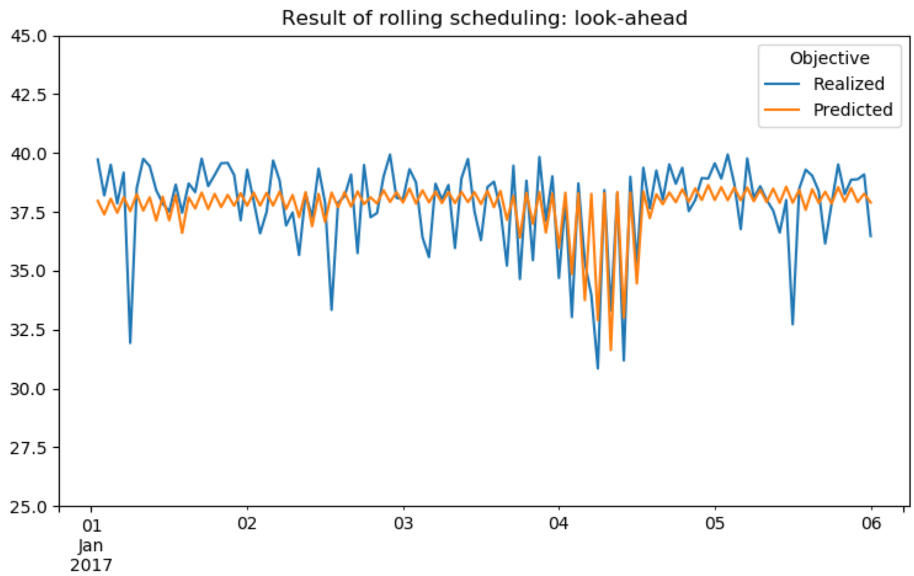


Figure 4.5: Predicted and realized objective function value from look-ahead scheduling

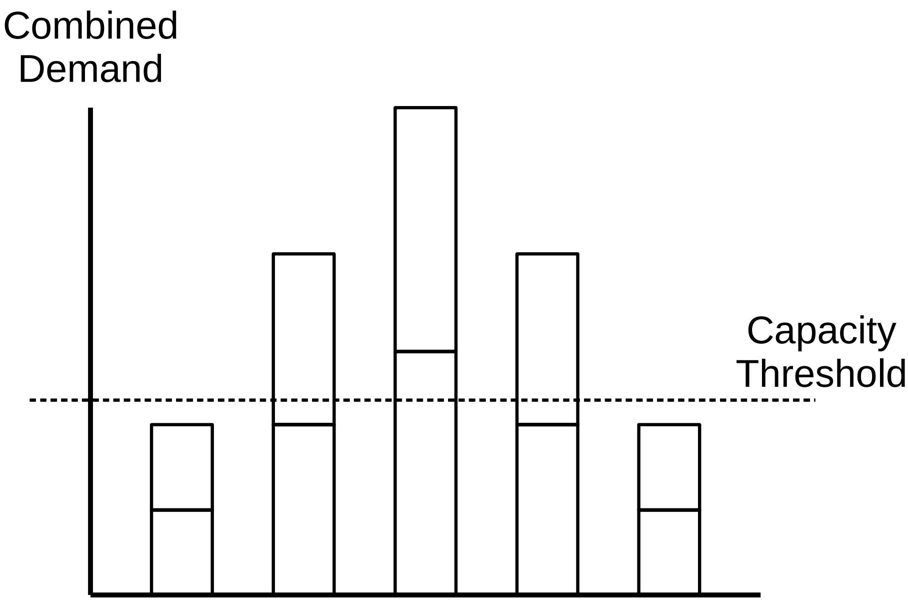


Figure 4.6: Demand on two feeders over 5 periods in a hypothetical case, Both feeders can be connected in period 1, only one feeder in periods 2 and 4, and neither feeder in period 3.

4.6.3 Some Observations on Myopic Planning

Notwithstanding the example in the previous section, myopic planning may be a robust heuristic for the Feeder Scheduling Problem, at least in the case where there are no supply quality constraints. The example in Section 4.6.2 was carefully constructed. The feeder rewards were selected to make the feeders with the fastest growth in variance appear to be less attractive in period 1 but not so low that selecting them in period 1 significantly reduces the objective in the lookahead policy.

Consider the following stylised case with 2 feeders over 5 periods illustrated in Figure 4.6. In period 1 both feeders are connected. Over the following 4 periods the deterministic mean will evolve in such a way that it is optimal to connect just 1 feeder in period 2, neither feeder in period 3, 1 feeder in period 4 and both feeders in period 5. Suppose that the variance about each feeder's demand level grows according to a concave function of the number of periods disconnected, $\sigma^2(t)$, which is the case if the feeder demand is given by an AR(1) process. This means there are diminishing marginal increases in uncertainty for every period the feeder is disconnected.

Suppose that the reward per unit demand on each feeder is the same and the size of the feeders is such that we are indifferent to which feeder we disconnect in period 2. The key decision to be made therefore is which feeder to reconnect first in period 4. We could connect the feeder that has been off for longest and about which we are most uncertain. i.e. the feeder disconnected in period 2. Alternatively, we could connect the feeder disconnected most recently about which we are most confident of its value, i.e. the feeder disconnected in period 3.

We will argue that the correct policy is to reconnect the feeder that has been disconnected for the least amount of time. This is the feeder that would be selected from a purely short-sighted perspective because the most recently disconnected feeder has lowest variance at period 4 and therefore minimizes the expected penalties in period 4. We now show that this strategy minimizes the expected penalties in period 4 and 5, so the myopic policy is in fact the correct policy.

Expected penalties are proportional to standard deviation of the total amount in the knapsack in each period. Therefore, it is important to keep the variance of the items used in the knapsack low. The standard deviation is the square root of the variance in each period, which is a concave function. The variance in each period is the sum of the variances of each item that is connected in that period.

The sum of the standard deviation if we reconnect the most recently disconnected feeder is:

$$\sqrt{\sigma^2(1)}+\sqrt{\sigma^2(0)+\sigma^2(3)}=\sqrt{\sigma^2(1)}+\sqrt{\sigma^2(0)+\sigma^2(2)+\delta}$$

The sum of the standard deviation if we reconnect the feeder that has been disconnected for longest is:

$$\sqrt{\sigma^2(2)}+\sqrt{\sigma^2(0)+\sigma^2(2)}=\sqrt{\sigma^2(1)+\varepsilon}+\sqrt{\sigma^2(0)+\sigma^2(2)}$$

Because of the concavity of the variance as a function of length disconnected $\sigma^2(2)-\sigma^2(1)=\varepsilon>\delta=\sigma^2(3)-\sigma^2(2)$

Now consider:

$\sqrt{\sigma^2(1)+\varepsilon}=\sqrt{\sigma^2(1)}+E$

$\sqrt{\sigma^2(0)+\sigma^2(2)+\delta}=\sqrt{\sigma^2(0)+\sigma^2(2)}+D$

The concavity of $\sqrt{\cdot}$ combined with the fact that $\varepsilon>\delta$ and $\sigma^2(1)<\sigma^2(0)+\sigma^2(2)$ means that $E>D$. In other words, the penalty for increasing the variance of the knapsack at period 4 from $\sigma^2(1)$ to $\sigma^2(2)$ is greater than the penalty of increasing the variance of the knapsack at period 5 from $\sigma^2(0)+\sigma^2(2)$ to $\sigma^2(0)+\sigma^2(3)$

This shows that in this case the myopic policy is better. In general, the picture is complicated by different stochastic processes for different feeders, by knapsack packing issues and by different reward levels per feeder. This is why in Section 4.6.2 we were able to give a case where myopic planning is not optimal. Nevertheless, we think the case we have just described gives some intuition for why myopic planning is a robust heuristic for the feeder scheduling problem. However, myopic planning would lead to items not being selected for long periods of time. It therefore conflicts with keeping disconnection intervals short.

4.6.4 Supply Quality Constraints

The results in Section 4.6.1 suggested that, without any supply quality constraints, DSOs can manage the risk of overconsumption and maximize revenue by planning which feeders to connect in a *myopic* way. We now investigate whether this is the case if the DSO must adhere to supply quality constraints.

If the DSO does not plan correctly they may find they have to connect too many customers in order to satisfy the supply quality constraints. They may also find that they must connect unattractive feeders at a worse time. This seems most likely to be an issue when the reward per unit demand varies by feeder. In this case the DSO may prefer to connect 'unattractive' feeders at times of relatively low demand so that at times of high demand they are free to use their capacity to supply more 'attractive' feeders. Planning correctly to avoid these outcomes presumably requires the DSO to take future periods into account.

We repeated the simulations from Section 4.6.1, imposing a constraint that each feeder can be disconnected for at most 3 periods. Table 4.4 shows the average realized objective value per period and the predicted objective value. The realized objective is not substantially different from the predicted objective and as in Section 4.6.1 there is a marginally larger difference with unequal feeder rewards.

It was not possible to find a solution to the 6 period lookahead problem with equal feeder rewards that was within optimality tolerance in reasonable time. We report the results of the best solution found allowing 10 minutes to find a solution at each time step. The expected solution substantially overestimates the actual objective. The actual objective of the solution is comparable to the solution for shorter horizons. This suggests that it is difficult to prove optimality. Possibly because of the very similar rewards per feeder it is difficult to get good bounds meaning that it is necessary to explore an excessively large proportion of the branch and cut tree to prove optimality.

Contrary to our expectation this example provides relatively little evidence of the benefits of planning ahead over myopic scheduling. The expected and actual objective of the 6 period lookahead is the best, but only marginally so. By contrast the 3 period lookahead is worse.

Nevertheless, we think that cases exist where failing to plan for future decision intervals can have very bad consequences if there are supply quality constraints. For example, consider a case where load is 3 times larger than supply and continues like that for several periods. Suppose that one third of the items are more attractive from a myopic perspective because they have a slightly lower variance or slightly higher value. Furthermore, suppose that the maximum outage length

Rewards	horizon obj	0	1	3	6
equal	Predicted	5808.6	5807.6	5786.0	6855.0
	Realized	5800.0	5802.7	5800.1	5827.5
unequal	Predicted	6196.1	6212.8	6087.0	6221.1
	Realized	6208.9	6206.3	6096.9	6221.9

Table 4.4: Average predicted and realized objective value (i.e. expected rewards minus expected penalties vs actual rewards minus actual penalties) for Central Path approximation of feeder scheduling problem with maximum outage length supply quality constraint. The approximate results from the problem that could not be solved to proven optimality are highlighted in red

is 2 hours. The myopic policy will be to choose the better value items for periods 1 and 2 and then must choose all the remaining items in period 3, resulting in a very large overconsumption penalty. A similar problem can arise with minimum supply frequency constraints.

Our results suggest that it is sufficient to solve the stochastic knapsack problem for the immediate decision interval but plan deterministically for subsequent decision intervals, not taking account of the effect of our decisions on future forecast variance.

4.7 Conclusions

We began this chapter by observing that DSOs face a continual problem of selecting which feeders to connect because the latent electricity demand in the DSO as a whole would exceed the available supply at all times. DSOs should prioritize certain feeders in order to maximize revenues but also manage the risk that the realized demand will exceed the supply that the SO allocates to them. We formulated this as the problem of maximizing the expected revenue from supplying energy minus the expected regulatory penalties should overconsumption occur. We gave a re-formulation of this MINLP problem as a more tractable MIP. This can be solved in reasonable time for problems of realistic size.

Our primary focus in this chapter was this issue of how the optimal scheduling policy is affected by the dependence of future forecast quality on current decisions. We gave a recursive formulation of this problem which can be tackled by a rolling horizon approach. In general, this category of problems is extremely difficult to solve. However, we noted that for a plausible model in which demand is modelled as an autoregressive process around a deterministic mean, modelling the effects of our decisions on the future forecast variance is a practical computation to perform in an optimization model.

In theory one should take into account the effect of one's decisions on the future state of knowledge. However, our computational experiments provide some evidence that the theoretical benefits of planning ahead are not realized in practice. It seems that for realistic range of stochastic demand processes a robust policy can be derived by a “myopic” approach. Where there are supply quality constraints our results suggest that it is sufficient to solve the stochastic knapsack problem for the immediate decision interval but plan deterministically for subsequent decision intervals, not taking account of the effect of decisions on future forecast variance.

Regulators may find this model useful to investigate the impact of different supply quality policies. We have already considered how minimum connection frequency and maximum outage length policies could be incorporated into the model. The model is flexible enough to be applied to cases with different supply quality constraints other than the ones we have considered.

Over longer forecast horizons it becomes increasingly implausible that the forecast errors are not correlated between feeders. Our results do not extend to this case because we cannot simply add the variance of the feeders. One way to include the correlations between feeders might be to formulate a model for the deterministic mean conditional on some exogenous covariates (temperature for example). Suppose for each feeder the mean demand is $\beta_i Y$ where Y is an exogenous covariate which has variance σ_Y^2 and β_i is the regression coefficient for feeder i. The total variance of the mean on all connected feeders is given by $\sigma_Y^2 \sum_i(\beta_i^2)$. A few features of this model may mean it is tractable. Firstly, the uncertainty about these exogenous covariates is not endogenous to the model so there is not additional complexity here. Secondly, the dimensionality of the additional uncertainty is only as great as the number of covariates. It may therefore be practical to use a scenario tree approach to integrate over the future evolution of the covariate. On the other hand, a quadratic term is required which makes the problem non-linear. Further study is required to see if this problem can be solved in reasonable time.

As we mentioned in the introduction, there is a benefit to being able to offer a predictable, if intermittent service to consumers. DSOs may wish to plan which feeders to connect and which feeders are ‘at-risk’ sufficiently far in advance that the the load shedding schedule can be communicated to customers. At the actual time of implementation of the load shedding schedule the DSO may need to revise this schedule in response to system conditions. Our work so far disregarded this in favour of investigating the case of scheduling disconnections in real time to manage overconsumption risk.

Our framework allows us to minimize the deviation from an existing schedule by imposing a cost for this. Figure 4.7 shows how the scheduling problem considered in this chapter can be thought of as the recourse problem of an upper level decision to set a provisional schedule.

Finding the optimal provisional schedule requires solving a complicated multi-stage mixed-integer stochastic program for which we do not propose a solution method here. However, it would be a simple step to apply our methods, not to short term forecasts based on the most recently available data, but instead to a model of the long run distribution of demand (and possibly capacity). The methods we have derived could then be used to determine which

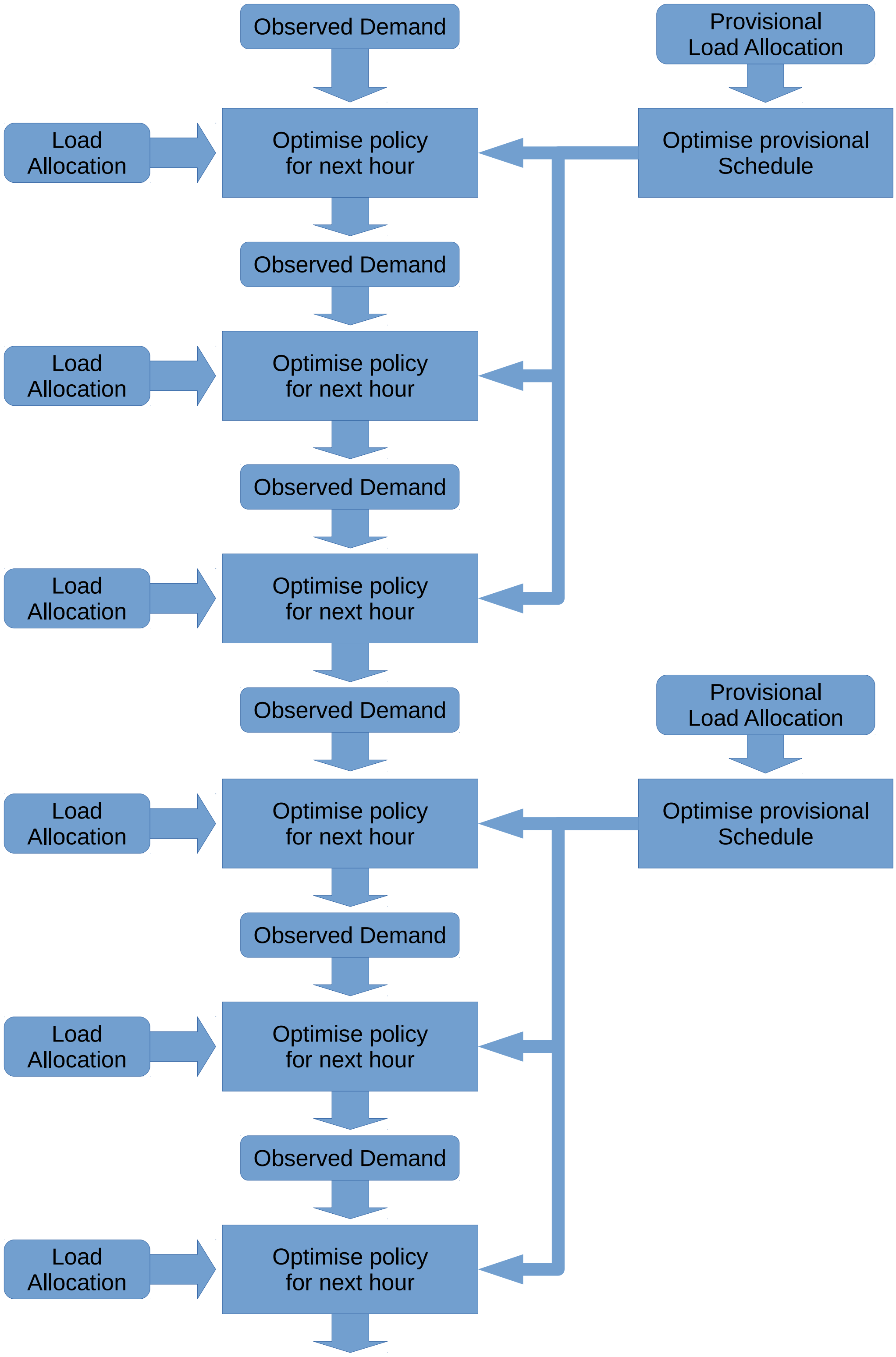


Figure 4.7: Schematic representation of the sequential nature of the feeder scheduling problem

additional feeders to connect if there is surplus capacity or which feeders to disconnect if demand exceeds capacity. Having been derived with long term, high variance forecasts this policy would likely be conservative. Therefore, although such a policy would not be optimal, it is unlikely that it would suffer large penalties for overconsumption or disconnecting customers that were not informed they were at-risk of load shedding.

An interesting area to extend our work would be to investigate whether we can derive a set of 'off the shelf' schedules which guarantee a given service level with reasonable confidence. The DSO could select a schedule based on prevailing supply/demand conditions and communicate this to customers.

We note some potential barriers to using this approach in practice.

Firstly, DSOs would need a system to reliably aggregate demand observations to make forecasts in real time and that they are able to disseminate instructions reliably. DSOs in many LEDCs are starved of funds to invest in advanced SCADA systems.

We have not been able to obtain any records of the SO's instructions to DSOs. We also lack billing and revenue collection data. Without this information it is not possibly to make a reasonable comparison between the methods we propose, and the actual scheduling system implemented by DSOs.

Furthermore we have assumed that revenue received is linearly proportional to load served to each feeder. It is possible that the willingness of customers to pay is affected by the level and timing of electricity supply they receive. It also may be the case that electricity supplied at different times is disproportionately consumed by customers in different tariff categories. All these factors should be subject to quantitative study.

We also assumed that electricity demand on each feeder is not affected by the pattern of load shedding – it is simply not supplied if the feeder is disconnected. In practice there is likely to be load shifting behavior that means the DSO's load shedding policy will affect demand in the short term. Over the longer-term customers might also react to changes in the predictability of their supply.

Given these factors, we suggest that DSOs should first develop a system of real time data collection and data validation. Mobile data collection systems that have been widely applied in public health may be of use here. Secondly, they should develop a system to produce demand forecasts and communicate these to local distribution SOs to feeder into their current ad-hoc decision-making process. This should be combined with reliable collection of data pertaining to billings, revenue collection, customer enumeration, equipment status and load allocation. It would then be possible to benchmark the current performance of the system and diagnose if better demand management has the potential to improve the utilization of allocated load.

We now look at the issue of power rationing from a higher level. In the next chapter we take the perspective of the system operator who must allocate scarce power between the needs of different electricity distribution companies.

Chapter 5

National Level Load Shedding

5.1 Introduction

In this chapter we take the perspective of the SO of a power system in a chronic power shortage. In power shortage conditions the SO must coordinate load shedding and dispatch generation so as to balance the connected demand with available power supply.

We analyse the case when the SO wants to dispatch generation and ration power supply in order to minimise load shedding while supplying different regions of the network with a fixed proportion of total power delivery. We have chosen to study this particular situation because there is a concrete example of this type of regulation implemented in Nigeria (see Section 5.3). Nevertheless, the methods we develop can be applied to other cases as we discuss in Section 5.4.

The existence of transmission constraints means that achieving the proportional targets exactly may not be compatible with minimising load shedding. The aim of this chapter is to develop methods to quantify the trade-off between these two objectives.

The material in this chapter is based on an article published in *Energy Policy* [27].

5.1.1 Existing Work

Sophisticated models have been proposed for efficient allocation of service interruptions in power shortage conditions in [39], [42], [16] and[7]. However, these rely on complex financial mechanisms which are unlikely to be practical in developing countries where institutional competence is often poor [1]. There is literature on optimal load shedding such as [21], [14] and [62]. However, these treat load shedding as an emergency measure to return the system to a steady state in the case of a contingency event. The contribution of our work is to consider the implications of the case when load shedding is a constant necessity for the operation of the system. [55] discusses how power rationing can be most effectively implemented to deal with long-term power shortages and the authors present case studies from developed and developing countries, including Chile, China, California, the Dominican Republic, Japan and Brazil. [53] attempts to draw lessons from the Californian experience of power rationing in 2001 that can be applied to developing countries. These are both empirical studies focusing on institutional and managerial issues. In contrast our work provides a decision support model targeted specifically at the needs of developing countries.

5.1.2 Our Approach

It is reasonable to assume that any SO will want to minimise the amount of load shedding, however the SO likely also has other objectives. For example, in the Nigerian context the MYTO stipulates that each DisCo should receive a fixed proportion of total energy delivered. We say that a DisCo that recieves less energy than its proportional target is in “shortfall” relative to the target. In the context of this chapter “shortfall” is the gap between the amount of power supplied in a region and the amount it should get if the proportional targets were respected. It does not mean the amount of unmet demand. We assume that the SO pursues an optimising rather than merely satisficing approach to this policy.

Finding an optimal policy is complicated by the physics of AC electricity transmission. The power injected and extracted from the power system flows according to Kirchoff’s laws. This power flow must be feasible with respect to the operating limits of the transmission system components. In many underdeveloped power systems the transmission network is highly constrained and operated close to its technical limits. It is important therefore to explicitly model the power flows in the transmission system.

A further complication is that the operating conditions of a power system vary within and between days. If the SO is concerned with optimising the distribution of energy delivered over a time horizon during which the operating conditions of the power system are expected to change then we must model this variation. One aspect of variation that is important in all power systems is the level of demand over the course of each day. In many developing world power systems the level and spatial distribution of available generation also varies between days because of generator faults or shortages of fuel. Furthermore, poor maintenance of generators and unreliable supplies of fuel mean that the level and spatial distribution of available generation is highly variable. The SO must take these into account in order to determine an optimal policy of load shedding and generation. We can model the variation in the operating conditions of the network by a set of operating points representing the range and relative frequency of different operating conditions. We can then optimise the dispatch of generation and demand over all operating points.

We propose that all these features can be captured by formulating the SO’s problem as a set of linked Optimal Power Flow (OPF) problems with dispatchable loads. Each problem represents a scenario of available generation and demand level.

5.1.3 Assumptions

In practice the SO does not have direct control of loads except except at the interface between the transmission and distribution systems. This means that the SO's direct control of load is limited to the ability to disconnect customers in large blocks which should only be done in extreme cases where it is necessary to protect the stability of the power system. Routine load shedding is therefore the responsibility of the DSOs who must shed load to respond to the SO instructions of how much power they are to consume at each grid supply point. We assume that the DSOs are obedient to the SO's instructions and that they can shed load in small quantities relative to the total supply at each grid supply point. The SO is assumed to be able to set a level of load shedding at each grid supply point as if it was a dispatchable load.

In practice the resultant load is erratic because: a) the DSOs have limited knowledge of the likely load; b) poor maintenance means that circuit components regularly fail; and c) the DSOs may have incentives not to behave as instructed. Nevertheless we think that useful questions can be answered by studying the system on the assumption of obedient DSOs.

This model also assumes the network is operated strictly within the rating of the equipment, but it may be that the SOs are able to flex these limits (e.g., overloading transformers etc.) in order to deal with the operational challenges faced. The model could be adapted by incorporating less restrictive constraints or using soft constraints.

We neglect issues of uncertainty about the future level of available generation and demand and instead assume that the uncertain future availability levels are represented by a set of operating points. This is a valid assumption if the SO is concerned with a policy that is optimal, on average, over the long term, and the distribution of the level of available generation is stationary. The assumption is also valid if the time-scale of the problem is sufficiently short, say a day, that only the level of unsuppressed demand is likely to change. The uncertainty about the level of demand over one day is relatively limited.

5.1.4 AC vs DC Power Flow Models

We can model the physics of power flows at varying levels of detail by using either the AC power flow equations or their DC linearization. The AC version more accurately simulates the power flows, including the effects of reactive power, transmission loses and line capacity limits. The advantage of the DC version is that the resulting OPF problem is a convex problem, meaning that it can be solved more quickly and to proven optimality. The DC power flow equations are widely used in techno-economic electricity models because they are a good approximation to the true behaviour of most power systems operating in normal conditions. However, for a power system operating close to its limits it is important to precisely model reactive power flows and transmission loses (see Section 5.3.3).

5.2 Methods

5.2.1 An Optimal Power Flow Model with Dispatchable Loads

For for a set of operating points of available generation and unsuppressed demand indexed by $t \in \mathcal{T}$ we can define an OPF problem to find the combination of load shedding at each grid supply point and power output from each generator that minimises load shedding. If we assume that the cost of load shedding is much higher than the cost of generation then we can omit precise modelling of generator costs and simply give each generator some nominal cost that is the same for each generator.

We can solve this OPF problem to find the pattern of generation and load shedding that requires the least load shedding. We refer to this as the *Minimum Load Shedding Problem*. The objective function to be minimised for the is given by equation (5.1). This is simply the negative sum of power supplied in all operating points. The constraints of The Minimum Load Shedding Problem are described by equations (5.2) to (5.14) using the sets, parameters and variables in Table 5.1. This formulation uses the AC power flow equations for each operating point defined using in polar coordinates as in [10]. We formulate the DC version by removing all reactive power constraints and variables from the problem and assuming that the voltage magnitude, v = 1 for all buses.

In addition to finding the minimum level of load shedding in each operating point, the Minimum Load Shedding Problem can be used to obtain an upper bound for power supply to each region. If each grid supply point is assigned to a region, we can set the cost of load shedding for load in that region to be higher than the cost of load shedding in other regions. This is preferable to simply removing the loads in other regions because it removes the possibility that there may be circumstances in which this might reduce the load that can be delivered to the region of interest.

The Minimum Load Shedding Problem is to:

$$\min -\sum_{r,t} p^{\mathrm{R}}_{r,t} \tag{5.1}$$

Table 5.1: Symbols used to define the OPF

Symbol	Description
	Sets
$\mathcal{G}$	Generators, indexed by g
$\mathcal{B}$	Buses, indexed by b
$\mathcal{B}$	Demands, indexed by d
$\mathcal{G}_b$	Generators at bus b
$\mathcal{D}_b$	Loads at bus b
$\mathcal{D}_r$	Loads in region r
$\mathcal{B}_b$	Buses connected to bus b
$\mathcal{R}$	Regions, indexed by r
$\mathcal{T}$	Operating points, indexed by t
b_0	slack bus
	Parameters
V_b^{UB}	Voltage magnitude upper bound at bus b
V_b^{LB}	Voltage magnitude lower bound at bus b
$P_{g,t}^{\mathrm{UB}}$	Real power generation upper bound for generator g
$P_{g,t}^{\mathrm{LB}}$	Real power generation lower bound for generator g
$Q_{g,t}^{\mathrm{UB}}$	Reactive power generation upper bound for generator g
$Q_{g,t}^{\mathrm{LB}}$	Reactive power generation lower bound for generator g
$P_{d,t}^{\mathrm{D}}$	Real power at load d
$Q_{d,t}^{\mathrm{D}}$	Reactive power at load d
$S_{bb'}^{\max}$	Apparent power limit on line bb'
$G_{bb'}$	Conductance of line bb'
$B_{bb'}$	Susceptance of line bb'
P_r	Proportional load allocation target for region r
L^s	Limit on total load allocation slack variables
W	Weighting parameter for shortfall minimisation objective
	Variables
$v_{b,t}$	Voltage magnitude at bus b, in operating point t
$p_{g,t}^{\mathrm{G}}$	Real power generation at generator g, in operating point t
$q_{g,t}^{\mathrm{G}}$	Reactive power generation at generator g, in operating point t
$\alpha_{d,t}$	Proportion of load shed at load d, in operating point t
$p_{d,t}^{\mathrm{D}}$	Real power consumed by load d in operating point t
$p_{r,t}^{\mathrm{R}}$	Real power consumed by loads in region r in operating point t
$q_{d,t}^{\mathrm{D}}$	Reactive power consumed by load d in operating point t
$P_{bb',t}^{\mathrm{L}}$	Real power on line bb' from bus b, in operating point t
$Q_{bb',t}^{\mathrm{L}}$	Reactive power on line bb' from bus b, in operating point t
θ_{bt}	Voltage phase angle at bus b, in operating point t
$d_{r,t}$	Target minus actual power supplied to region r at operating point t
$s_{r,t}$	Shortfall from load allocation in region r at operating point t
s_r^{L}	Long-term shortfall from load allocation in region r

subject to

$$\sum_{g\in\mathcal{G}_b} p^{\mathrm{G}}_{g,t} = \sum_{d\in\mathcal{D}_b} p^{\mathrm{D}}_{d,t} + \sum_{b'\in\mathcal{B}_b} P^{\mathrm{L}}_{bb',t}, \tag{5.2}$$

$$\sum_{g\in\mathcal{G}_b} q^{\mathrm{G}}_{g,t} = \sum_{d\in\mathcal{D}_b} q^{\mathrm{D}}_{d,t} + \sum_{b'\in\mathcal{B}_b} Q^{\mathrm{L}}_{bb',t}, \tag{5.3}$$

$$p^{\mathrm{D}}_{d,t} = (1-\alpha_{d,t})P^{\mathrm{D}}_{d,t}, \tag{5.4}$$

$$q^{\mathrm{D}}_{d,t} = (1-\alpha_{d,t})Q^{\mathrm{D}}_{d,t}, \tag{5.5}$$

$$0 \le \alpha_{d,t} \le 1, \tag{5.6}$$

$$p^{\mathrm{R}}_{r,t} = \sum_{d\in\mathcal{D}_r} p^{\mathrm{D}}_{d,t}, \tag{5.7}$$

$$\begin{aligned} P^{\mathrm{L}}_{bb',t} = v^2_{b,t} g_{bb} + v_b v_{b'} (g_{bb'} \cos(\theta_{b,t} - \theta_{b',t}) \\ + b_{bb'} \sin(\theta_{b,t} - \theta_{b',t})), \end{aligned} \tag{5.8}$$

$$\begin{aligned} Q^{\mathrm{L}}_{bb',t} = -v^2_{b,t} b_{bb} + v_b v_{b'} (g_{bb'} \sin(\theta_{b,t} - \theta_{b',t}) \\ - b_{bb'} \cos(\theta_{b,t} - \theta_{b',t})), \end{aligned} \tag{5.9}$$

$$\theta_{b_0} = 0, \tag{5.10}$$

$$V^{\mathrm{LB}}_b \le v_{b,t} \le V^{\mathrm{UB}}_b, \tag{5.11}$$

$$P^{\mathrm{LB}}_g \le p_{g,t} \le P^{\mathrm{UB}}_g, \tag{5.12}$$

$$Q^{\mathrm{LB}}_g \le q_{g,t} \le Q^{\mathrm{UB}}_g, \tag{5.13}$$

$${P^{\mathrm{L}}_{bb',t}}^2 + {Q^{\mathrm{L}}_{bb',t}}^2 \le (S^{\max}_{bb'})^2 \tag{5.14}$$

Equations (5.2)–(5.3) are Kirchhoff's Current Law (KCL), enforcing real and reactive power balance, (5.4)–(5.6) define the power consumption in terms of the proportion $\alpha_{d,t}$ of real and reactive load shed at each demand bus d and operating point t, (5.8)–(5.9) are Kirchhoff's Voltage Law (KVL), (5.10) removes the degeneracy in the bus voltage angles by fixing it to zero at the arbitrary reference bus, (5.11)–(5.13) are constraints on voltage and power generation and (5.14) is the line flow constraints.

The line conductance $g_{bb'}$ and susceptance $b_{bb'}$ are defined by

$$g_{bb'} = \frac{r_{bb'}}{r^2_{bb'} + x^2_{bb'}}, \qquad b_{bb'} = \frac{-x_{bb'}}{r^2_{bb'} + x^2_{bb'}},$$

where $r_{bb'}$, $x_{bb'}$ are the line resistance and reactance, and parameters $G_{bb'}$ and $B_{bb'}$ are defined by

$$g_{bb'} = -\tau_{bb'} G_{bb'} = -\tau_{bb'} G_{b'b} = G_{b'b'} = \tau_{bb'}^2 G_{bb}, \tag{5.15}$$

$$b_{bb'} + 0.5 b_{bb'}^{\mathrm{C}} = B_{b'b'} = \tau_{bb'}^2 B_{bb}, \tag{5.16}$$

$$-b_{bb'} = \tau_{bb'} B_{b'b} = \tau_{bb'} B_{bb'}, \tag{5.17}$$

where $b_{bb'}^{\mathrm{C}}$ is the line charging susceptance and $\tau_{bb'} = 1$ except in transformer 'lines', where it is the tap ratio and (as in the MATPOWER~[63] convention) the ideal transformer is at the b end of the line.

Solution Method

The problem defined above is trivially decomposable. We can see this because none of the constraints involve more than a single operating point, t and the objective function is a sum over all operating points. We solve each problem using the Interior Point Method implemented in IPOPT [60]. The AC OPF is a non-linear, non-convex problem so there is no theoretical guarantee that the solver will converge to the globally optimal solution. In addition to converging to a locally optimal solution, the solver could converge to a stationary but non-optimal point or converge to an infeasible point. However, methods that guarantee global optimality are not yet a practical approach for the size and structure of problems we consider (a meshed network of 637 buses in our case study). In practical experience, local optima of the AC OPF are relatively rare [10]. Nevertheless, we address the risk of finding a local optimum by solving each cases from different initial points. For the first run we set $v_b = 1$ and $\delta_b = 0$ for all buses b, and $p_{g,t}$ and $q_{g,t}$ at the midpoint of their upper and lower bounds for each generator g. For subsequent runs the values of these decision variables were randomised. Using this method we have not been able to identify any local optima in the problem we consider in the case study.

5.2.2 The Short-Term Load Allocation Problem

In the problem described above we just minimise the load shed. We now show how to incorporate requirements for proportional fairness into the model. We can model these requirements easily by assigning each grid supply point to a region. For each region and for each operating point we impose a constraint that the sum of load supplied to that region must be greater than or equal to a given proportion of the total load supplied in all areas. To avoid computational issues caused by the difficulty of satisfying this constraint exactly we impose this as a soft constraint by penalising total violation of this constraint in the objective function.

In this formulation the proportional constraints must be satisfied in each operating point so regions in shortfall with respect to the target in one operating point cannot be compensated for this by a surplus in other operating points. This models the situation where the SO is interested in achieving the desired balance between proportional fairness and minimising load shedding over a short time horizon during which the operating conditions of the power system are expected to remain constant. Therefore, we call this the *Short-term Load Allocation Problem*.

The short-term load allocation problem can be used to quantify the trade-off between maximising power supplied and minimising the deviation from the load allocation targets. By varying the size of the penalty for deviation from the load allocation we can find various points on the Pareto frontier of the two objectives. By considering just a single operating point we can obtain the Pareto frontier for that specific scenario of demand and supply conditions. Alternatively, we can find the optimal trade off over a set of scenarios that places a consistent value on the relative importance of the two objectives.

The objective function to be minimised for the Short-term Problem is given by equation (5.18). This is the weighted sum of a penalty for violating the load allocation constraints in every operating point minus the sum of power supplied in all operating points. Constraint (5.19) defines the difference, $d_{r,t}$, between the power supplied in region r at operating point t and the corresponding target power supply for the region. The target power supply for region r is given by the total power supply to all regions multiplied by P_r, the proportion of power that should flow to region r according to the load allocation regulation. Recalling the definition of shortfall given in Section 5.1.2, the shortfall in region r at operating point t, $s_{r,t}$, defined in constraint (5.20), is the positive part of $d_{r,t}$.

The Short-term Load Allocation Problem is to:

$$\min W \sum_{r,t} s_{r,t} - \sum_{r,t} p^{\mathrm{R}}_{r,t} \tag{5.18}$$

subject to constraints (5.2) to (5.14) and also subject to

$$d_{r,t} = P_r \sum_{r'} p^{\mathrm{R}}_{r',t} - p^{\mathrm{R}}_{r,t} \tag{5.19}$$

$$\left.\begin{aligned} s_{r,t} &\geq 0 \\ s_{r,t} &\geq d_{r,t} \end{aligned}\right\} \tag{5.20}$$

Short-term Load Allocation Problem Solution Method

The Short-term Problem is separable by operating points. Each operating point defines an OPF which is independent from the others. We can see this by reformulating (5.18) as:

$$\sum_t \min W \sum_r s_{r,t} - \sum_r p^{\mathrm{R}}_{r,t} \tag{5.21}$$

and noting that none of the constraints include more than one value of t.

We can obtain the Pareto frontier of the two objectives by varying the value of W. The downside of this approach is that it does not guarantee a good spread of points on the Pareto frontier and cannot obtain points on concave parts of the Pareto frontier if they exist.

In the case where we just have a single operating point rather than a set of scenarios we can use the ε-constraint method described in [59]. This involves optimising just with respect to the power supply objective and imposing a constraint on the total shortfall. In this approach we set W to a very small positive value so that the solver will prefer solutions with a smaller level of shortfall, all other things being equal. This approach is not directly applicable to the case with multiple scenarios because constraining the total shortfall would prevent us from decomposing the problems by operating points. The Lagrangian decomposition scheme which we develop in section 5.2.3 for another purpose could be applied to this problem.

5.2.3 The Long-Term Load Allocation Problem

Suppose the SO is interested in achieving the desired balance between proportional fairness and minimising load shedding over a time horizon during which the operating conditions of the power system are expected to change. This means that shortfall in one operating point can be compensated by a surplus in another operating point. We refer to this as the *Long-term Load Allocation Problem*.

The objective function to be minimised for the Long-term Problem is given by Equation (5.22). This is the weighted sum of a penalty for violating the load allocation constraints considering all operating points together minus the sum of power supplied in all operating points. The long-term shortfall, s^L_r defined in constraint (5.23) is the positive part of the difference terms summed over all operating points t. It is defined for every region. The difference with (5.20) is that in (5.23) is that difference terms $d_{r,t}$ are summed over all operating points which has the required effect that shortfall in one operating point can cancel out surplus in another.

The Long-term Load Allocation Problem is to:

$$\min W \sum_r s^{\mathrm{L}}_r - \sum_{r,t} p^{\mathrm{R}}_{r,t} \tag{5.22}$$

subject to constraints (5.2) to (5.14) and instead of (5.20) subject to

$$\left.\begin{aligned} s_r^{\mathrm{L}} &\geq 0 \\ s_r^{\mathrm{L}} &\geq \sum_t d_{r,t} \end{aligned}\right\} \tag{5.23}$$

Long-term Load Allocation Problem Solution Method

The Long-term Load Allocation Problem becomes intractable for large numbers of operating points but unlike the short-term Load Allocation Problem it is not separable by operating points because of the complicating constraint (5.23). We therefore use an approach based on Lagrangian relaxation to decompose the problem. Because of the non-convexity of the original problem this approach is to some extent heuristic. However, as we show in Section 5.3, we can solve problems to a very small duality gap using this method. Lagrangian relaxation is applied to (5.23) (retaining the positivity constraint $s_r^{\mathrm{L}} \geq 0$) and the objective (5.22) to get the Lagrangian function (with weighting parameter W determining the balance between the two primal objective terms):

$$\begin{gathered} L(\boldsymbol{\mu}, W) = \min W \sum_r s_r^{\mathrm{L}} - \sum_{r,t} p_{r,t}^{\mathrm{R}} - \sum_r \mu_r (s_r^{\mathrm{L}} - d_{r,t}) = \\ \min W \sum_r s_r^{\mathrm{L}} - \sum_{r,t} p_{r,t}^{\mathrm{R}} - \sum_r \mu_r (s_r^{\mathrm{L}} - P_r \sum_{r',t} p_{r't}^{\mathrm{R}} + \sum_t p_{r,t}^{\mathrm{R}}) \end{gathered} \tag{5.24}$$

subject to constraints (5.2) to (5.19) and $s_r^{\mathrm{L}} \geq 0$.

By the weak duality theorem this gives a lower bound on the primal problem for any choice of $\boldsymbol{\mu} \geq 0$. Furthermore, this can be rearranged to:

$$L(\boldsymbol{\mu}, W) = \sum_r \min s_r^{\mathrm{L}} (W - \mu_r) + \sum_t L_t(\boldsymbol{\mu}, W) \tag{5.25}$$

where,

$$L_t(\boldsymbol{\mu}, W) = \min \sum_r p_{r,t}^{\mathrm{R}} (-1 + \sum_{r'} (\mu_{r'} P_{r'}) - \mu_r) \tag{5.26}$$

Therefore (5.24) is separable by operating points because the none of the constraints mention more than one value of t

For a given $\boldsymbol{\mu}$ we can solve $L(\boldsymbol{\mu}, W)$ by solving $|\mathcal{T}|$ independent OPF problems $L_t(\boldsymbol{\mu}, W)$, where $\mathcal{T}$ is the set of operating points, and calculating the optimal $s_r^L \geq 0$ to minimise

$$\sum_r \min s_r^{\mathrm{L}}(W - \mu_r)$$

The best lower bound will be given by $\max_{\boldsymbol{\mu}} L(\boldsymbol{\mu}, W)$, from which we can observe that the optimal $\boldsymbol{\mu} \in [0, W]$, otherwise $L(\boldsymbol{\mu}, W)$ is unbounded below.

To find the values for $\boldsymbol{\mu}$ which maximise $L(\boldsymbol{\mu}, W)$ we can build a cutting plane approximation to $L(\boldsymbol{\mu}, W)$. From the solution of $L(\boldsymbol{\mu}, W)$ for any particular $\boldsymbol{\mu}$ we can compute the gradient$\nabla_{\boldsymbol{\mu}} L(\boldsymbol{\mu}, w)$, from $\frac{\partial}{\partial \mu_r} L(\boldsymbol{\mu}, w) = \hat{s}_r - P_r \sum_{r'} \hat{p}^{\mathrm{R}}_{r',t} + \sum_t \hat{p}^{\mathrm{R}}_{r,t}$, which gives us the cutting plane $f \leq L(\boldsymbol{\mu}, W) + (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \nabla_{\boldsymbol{\mu}} L(\boldsymbol{\mu}, W)$. For a set of cutting planes $\mathbb{P}$ we solve the master problem $\max_{f \in \mathbb{R}} f$ subject to $\mathbb{P}$ and $0 \leq \hat{\mu}_r \leq W$ to obtain an updated $\hat{\boldsymbol{\mu}}$. In practice we find that a dual stabilisation method is needed for the dual problem to converge in reasonable time. We use a Proximal Bundle Method (see [28] for example).

5.3 Case Study: Nigeria

Nigeria exhibits the features that motivate our modelling approach: demand exceeds the power system's ability to supply power and the Nigerian regulator NERC has adopted an approach to manage the regional rationing of power based on proportional Load Allocation targets for each distribution company.

As a result of the recent privatization and restructuring of the Nigerian power system there are 11 private utility companies with regional distribution and retail monopolies known as DisCos. The roles of the transmission service provider and SO are combined in the government-owned TCN. Generation is owned by privatized GenCos but centrally dispatched by TCN and the DisCos are required to pay the GenCos for energy received according to generation tariffs set by the NERC.

The proportional load allocation targets set by NERC are shown in Table 5.3. These roughly correspond to the proportion of customers in each DisCo. The precise implementation of the load allocation policy has varied since privatization, but the principle is that load supplied in each DisCo region over some time period should be a fixed proportion of the total load supplied.

In this case study we analyse how enforcing this regulation over different time scales and with varying levels of strictness affects the total amount of power that can be delivered. Requiring the regulation to be satisfied over a short time horizon during which the operating conditions of the power system are expected to remain constant corresponds to the Short-term Load Allocation Problem. Requiring the regulation to be satisfied on average, over the long term corresponds to the Long-term Load Allocation Problem. In order to explore the effect of enforcing the policy with varying levels of strictness we obtain the Pareto frontier of the two objectives: minimising the total load shed and minimising a measure of the deviation from the proportional targets in all regions that have a shortfall. We first consider each of the operating points individually and obtain the Pareto frontier of power supply and shortfall in each operating point. We then obtain Pareto frontier of both the Long-term and Short-term Load Allocation problems for this set of operating points.

5.3.1 Network Topology

The Nigerian transmission system consists of 330KV and 132 KV lines as shown in figure 5.1. The transmission system topography consists of a meshed grid in the south west of the country connected to the north and east by radial lines.

We consider a 637 bus single phase equivalent circuit model of the Nigerian transmission system[1]. The model contains data on the generation assets and the 330KV and 132KV transmission system. 111 buses are at the 330KV level, 176 at the 132KV level and 190 at the 33KV level. The remainder are low voltage load buses or generator buses. All the branches connecting to buses at a lower voltage than 132KV are transformers, not transmission lines. Loads are aggregated at grid supply points where the voltage is stepped down to distribution level. The model is intended to represent the state of the transmission network at the end of 2014.

5.3.2 Approximating the Variation in Network Operating Conditions

Our problem formulation is general enough that the set of operating points can represent any variation in the operating conditions of the network including generation availability, demand and transmission system topology. In the Nigerian case it is the variation in the level of available generation that is the major cause of supply variability. Variation in the network topology due to faults is of a lower order of significance.

Generation

The level and distribution of generation available is exogenous to our model and it is highly variable. Approximately 80% of the on-grid generation capacity of the Nigerian system is supplied by 18 gas fired power stations. These are all located in the south of the country. The remainder is supplied by 3 hydro-electric power stations located on the Niger river in the central western part of the country. The theoretical maximum capacity at all power generating sites connected to the national grid, as shown in table 5.2, is 11,345 MW. However, almost half of this is unavailable due to plant maintenance, unreliable gas supply and long term plant outages.

We have obtained daily reports produced by TCN detailing the operations of the power system over 1 year from May 2014. Figure 5.2 shows total available capacity in this period which

1. This model was provided and validated by EMRC http://energy-mrc.co.uk/

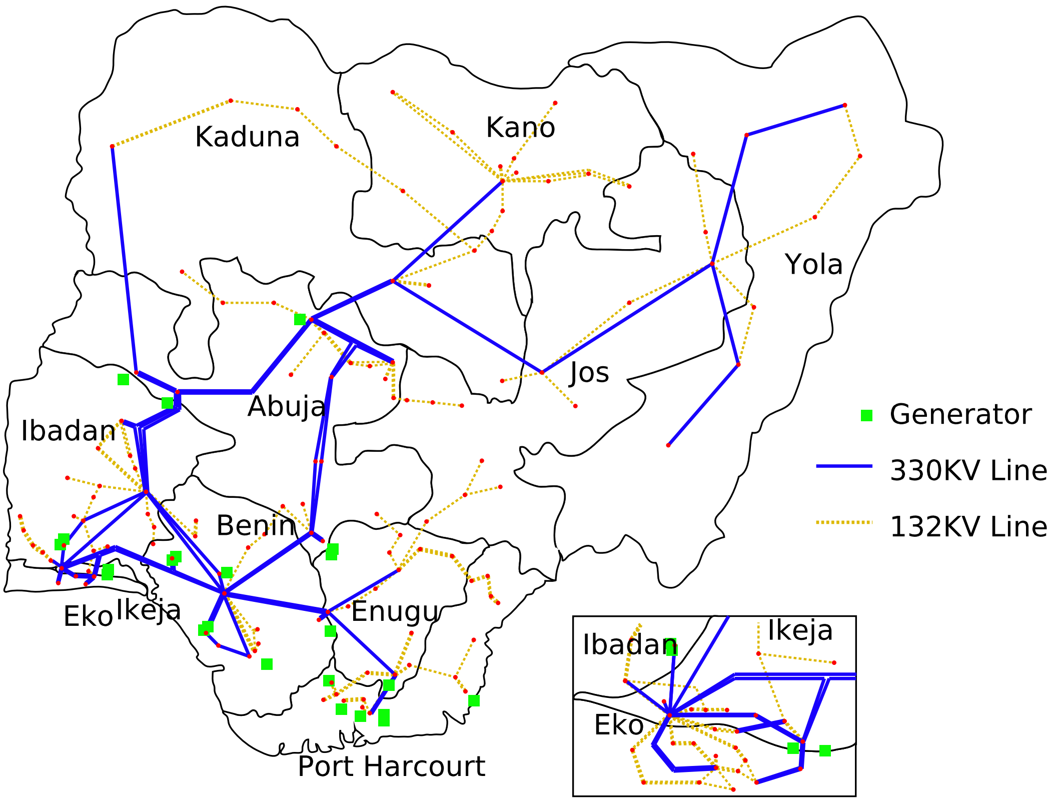


Figure 5.1: Map of Nigeria showing distribution company areas, high voltage transmission lines and location of generators

Table 5.2: On-Grid generation capacity April 2014 to January 2016

Generator Name	Nameplate Capacity (MW)	Mean Installed Available Capacity (MW)
Thermal Generators		
AES	270	98
Afam IV-V	600	34
Afam VI	650	450
Delta	972	328
Egbin	1320	733
Geregu	414	145
Geregu NIPP	434	170
Ibom Power	190	51
Ihovbor	450	168
Okpai	450	398
Olorunsogo	335	140
Olorunsogo NIPP	675	169
Omoku	250	0
Omotosho	335	126
Omotosho NIPP	450	198
Rivers	150	94
Sapele	1020	70
Sapele NIPP	450	137
Hydro Generators		
Jebba	570	321
Kainji	760	142
Shiroro	600	309
On-Grid Total	11345	4281

can be seen to be highly variable. The peak daily power output from all these plants over the study period was on average 3927 (The period of peak generation is not at the same time every day and does not correspond to the period of highest demand because demand always exceeds supply). Peak generation quite closely tracks available capacity which indicates that the system is often constrained by the level of available generation

We created 26 operating points with total available capacity evenly spaced from 3 to 5.5 GW in increments of 100 MW by scaling the nameplate real power generation capacity of each generator by the same factor so that the spatial distribution of generation is the same.

We also create a set of 21 generation operating points that implicitly approximate the probability distribution of the level and spatial distribution of available generation. For our study we selected a set of 21 days uniformly spaced over the year starting May 2014 and create an operating point to match generation availability on that day by scaling the maximum generation output of each generators to match its availability on that day. 21 operating points were chosen to balance the goals of minimizing computational complexity while approximating the space of past variation reasonably well.

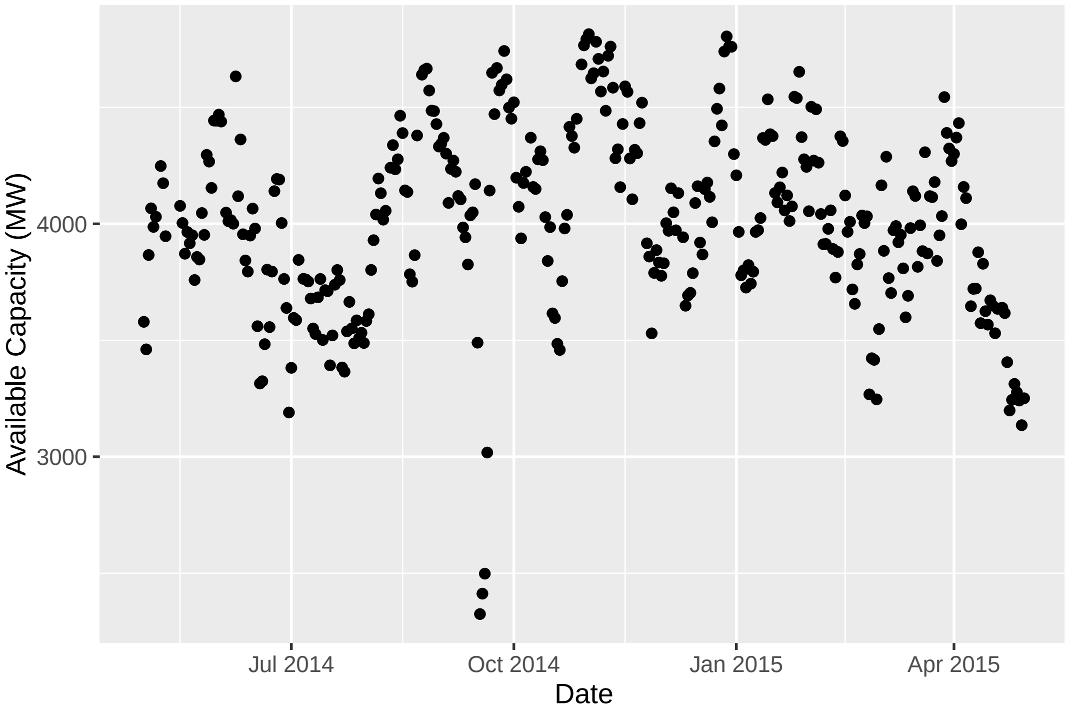


Figure 5.2: Time series of available generation capacity

Loads

As we have observed, the Nigerian power supply system fails to meet total electricity demand. The load which is served at any given moment is referred to as the suppressed load. The loads which would consume power but are not currently supplied are referred to as the shed load. The combination of the suppressed load and the shed load is the unsuppressed load. The shed load cannot be precisely measured so there is uncertainty about the unsuppressed load.

Because load shedding is endogenous to our model, we require that the loads included in the model reflect the unsuppressed load at each grid supply point.

We have obtained one set of measurements of actual connected load. In the absence of more data we assume that the measurements are representative of the power factor at grid supply points and the relative scale of the loads within distribution company regions. The measurements do not include the shed load, however, so we scale the measurements of connected load up maintaining the same power factor and ensuring that the total load in each DisCo is proportional to it's load allocation target.

There is substantial uncertainty about the level of unsuppressed demand. The TCN daily reports for the year starting May 2014 contain estimates of the peak load from 13,070 and 14,630 MW. At the much lower end the 2009 Tractabel study forecast the peak load in 2015 to be between 5500 MW and 7500 MW. Neither of these estimates capture demand variation over the course of the day. An alternative approach would be to apply our dynamic regression model as described in Section 3.5. We could then include operating points reflecting a range of peak, off-peak and mid-peak demand conditions.

However, if demand is sufficiently high that transmission and generation constraints always prevent the load being fully supplied in any DisCo then it may not be so important to include an accurate level of load in the model as long as it is high enough. Our approach is to scale the level of total load to 10 GW for our analysis and carry our a sensitivity analysis of these results in Section 5.3.3.

5.3.3 Results

Minimum Load Shedding

We found the minimum load shed for the 26 operating points with total available capacity evenly spaced from 3 to 5.5 GW in increments of 100 MW. The load supplied in each operating point is show by the green diamonds in Figure 5.3.

We see that level of power generation is highly dependent on level of available generation for scenarios of available generation up to about 4.8GW. After this point increases in generator availability don't translate to proportionate increases in power supplied even though there is still substantial load shedding. This implies that without reinforcement of the transmission network, improvements in generation availability will not be able to improve the power supply beyond this limit.

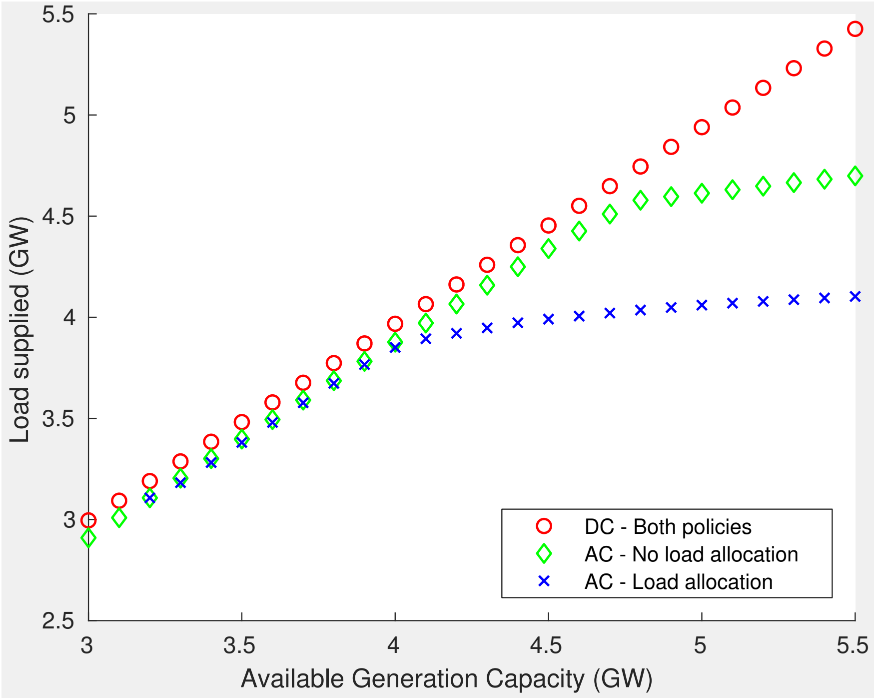


Figure 5.3: Variation in supplied load with changes in generation availability under AC and DC models

Table 5.3 shows the proportion of load supplied that goes to each DisCo in the solution to the minimum load shedding problem for 6 of the operating points. The most substantial undersupply is to Eko and Yola. While Eko's load share increases with more available generation the proportion of power going to Yola falls. Eko is in the South West, where there is a concentration of generators, but also a concentration of other DisCos with relatively large loads. Eko loses out in low available capacity scenarios because it competes with the other DisCos in the region for power. When there is higher available capacity it can absorb more power. On the other hand, Yola is in the North East, far from generators, at the end of long transmission lines. In higher generation scenarios, transmission constraints stop it from absorbing more power.

Feasibility and Cost of Load Allocation Policies

We now turn our attention to analysing how the transmission system constrains the SO's attempts to meet their load allocation policies. When we minimize load shedding it is possible that relatively small differences in the transmission losses to supply different regions cause substantial shortfalls for some DisCos. If this is the case then it should be possible to meet the load allocation targets with a relatively small increase in load shedding.

Table 5.3: Distribution of load under a policy of maximising power supply in 6 generation scenarios

Gen capacity	Load supplied	Abuja	Benin	Eko	Enugu	Ibadan	Ikeja	Jos	Kaduna	Kano	Port Harcourt	Yola
3000	2910	12.1%	11.9%	5.7%	10.2%	12.0%	9.9%	6.3%	16.3%	8.0%	6.5%	1.1%
3500	3398	10.3%	9.0%	9.1%	13.5%	13.0%	10.1%	5.5%	14.1%	6.9%	7.6%	1.0%
4000	3877	10.7%	10.4%	10.0%	13.4%	12.1%	12.2%	5.5%	11.4%	5.7%	7.6%	1.0%
4500	4339	11.9%	10.0%	8.1%	12.4%	12.5%	15.3%	5.5%	11.1%	5.1%	7.6%	0.6%
5000	4607	12.5%	9.3%	7.3%	11.7%	12.7%	15.0%	5.4%	11.1%	5.6%	8.2%	1.3%
5500	4692	12.2%	9.1%	7.6%	11.9%	12.6%	15.0%	5.4%	11.0%	5.7%	8.5%	1.2%
MYTO2 Load Allocation		11.5%	9.0%	11%	9%	13%	15%	5.5%	8%	8%	6.5%	3.5%

The blue crosses in Figure 5.3 show the total load supplied when we enforce the load allocation policy in the 26 scenarios of evenly spaced available generation between 3 and 5.5 GW. For comparison recall that the green diamonds show the load supplied when we minimize load shedding without taking the proportional targets into account. There is a substantial reduction in load supplied if the load allocation targets are applied when the level of available generation exceeds 4 GW.

Comparison of the AC and DC Models

We considered carrying out our analysis with the AC and DC versions of the models. Here we present the results that motivate our decision to use the AC model. We repeated the model runs of Section 5.3.3 and Section 5.3.3 using the DC version of the model. Figure 5.3 shows the total supplied load plotted against the available generation capacity.

More load is supplied in the solution to the DC model than is possible with the AC model, particularly for higher generation scenarios. Furthermore, in the DC model the load allocation policy has no effect on delivered load. We can see from the large difference in the results for the AC and DC models that accurate modelling of losses and reactive power flows is important to get accurate results for the Nigerian system. These results suggest that there is a shortage of reactive power in the network and that installation of reactive compensation could help to alleviate some of the transmission constraints.

Based on these results we carry out our analysis with the AC model. The DC approach could be modified by adding some approximation of the losses (e.g. a constant loss model) and using a slightly lower line flow limit to account for reactive power flows. However, it is unlikely that any approximation of losses and reactive power flows would give accurate results in the wide range of different operating conditions of the network that we explore.

An Illustrative Example of Change in Regional Distribution of Power with Increasing Strictness of Load Allocation Targets

Recall that shortfall, as defined in constraint (5.20), is the gap between the amount of power supplied in a region and the amount it should get if the proportional targets were respected. There are essentially three ways to reduce the regional shortfall. First, it may be possible to simply reallocate load from one DisCo to another, which reduces overall power supply only as much as the transmission losses are increased. Second, if the network is more congested, load can be reallocated in such a way that it not only increases the transmission losses but impedes supply to other DisCo. Lastly, in the extreme case where it is not possible to increase the supply to under-supplied DisCos, we can arbitrarily disconnect loads in well supplied DisCos until the load supplied in all DisCos is the correct proportion of load supplied.

We illustrate this in the following plots. We studied the scenario with 4500 MW of available generation capacity. Starting from a case in which we found the minimum load shedding solution we gradually increased the weight given to the objective of minimizing deviation from the proportional targets.

The resulting reduction in regional shortfall as we increase the weight parameter has 3 distinct phases. In the first phase deviation from the targets is reduced from 400 MW to 100 MW with relatively little decrease in overall load supplied. The regional distribution of load that corresponds to this is shown in figure 5.6. Each horizontal line represents the solution to the optimization problem with the constraint on deviation below the regional load allocation targets set at a particular level. As this level changes, the distribution changes. This shows that in order to reduce the deviation from 400MW to 100 MW, power supply is reduced in Kaduna DisCo and increased in Kano. These are 2 DisCos without local generation and supplied by the same radial transmission lines from the central network. This implies there is a common limit on how much can be transferred to these DisCos. Reallocating load within this DisCo reduces overall power supply only as much as the transmission losses are increased.

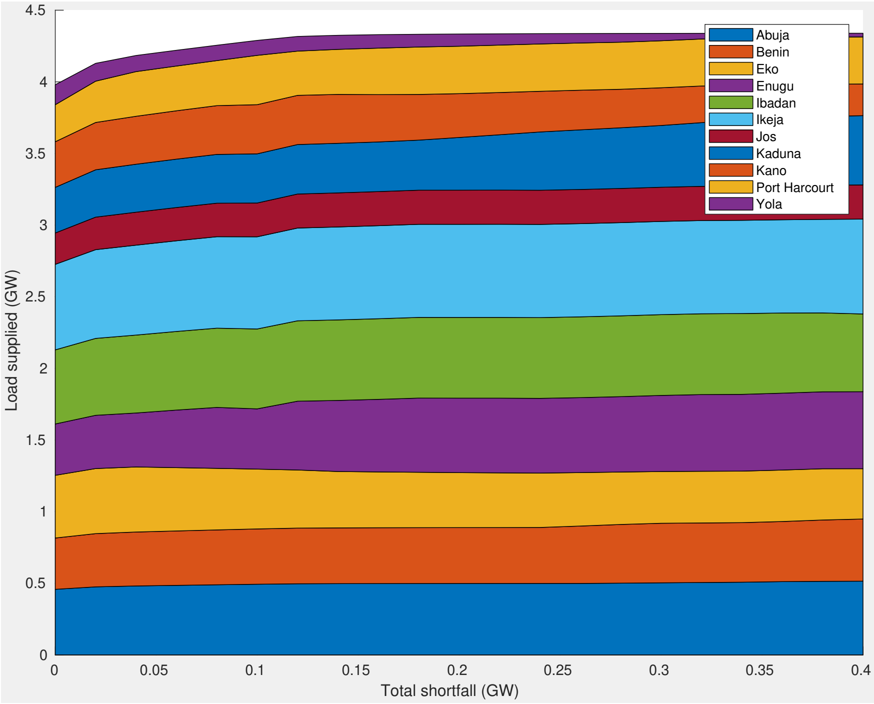


Figure 5.4: Total load supplied, split by DisCo as allowable deviation from load allocation targets is reduced from .4GW to 0GW

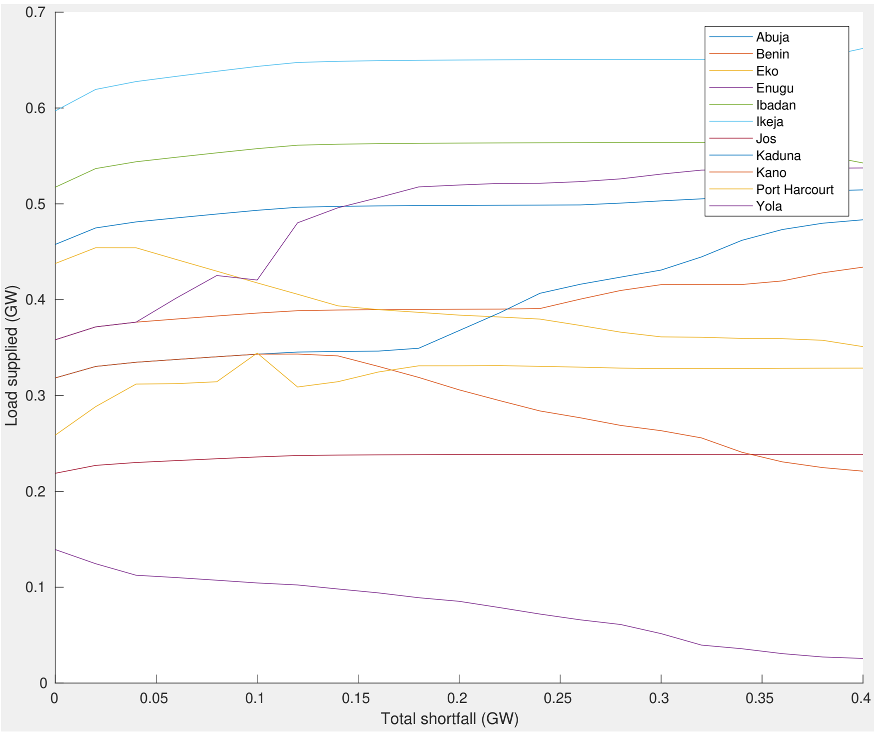


Figure 5.5: Load supplied to each DisCo as allowable deviation from load allocation targets is reduced from .4GW to 0GW

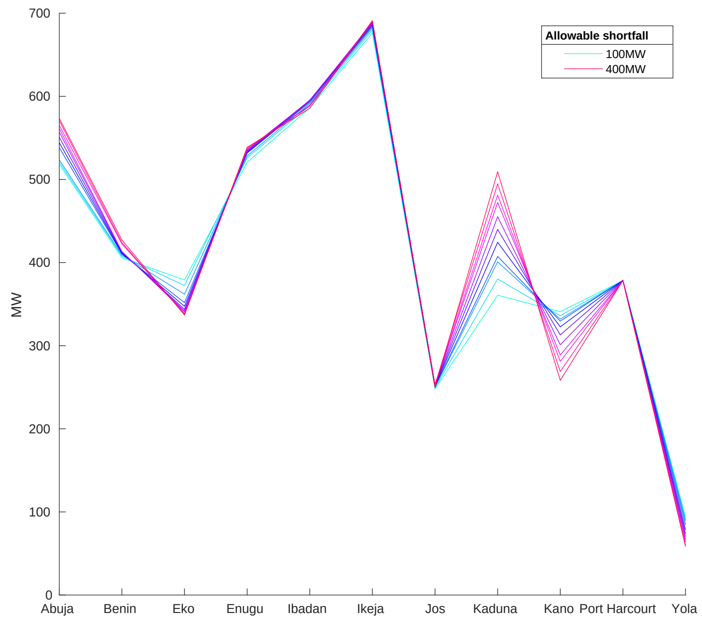


Figure 5.6: Change in regional distribution of load as allowable deviation from load allocation targets is reduced from 400MW to 100MW

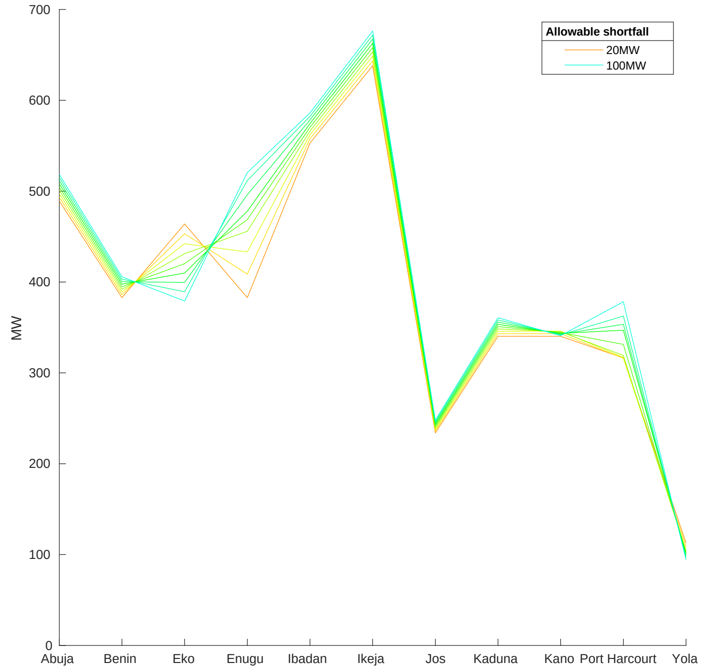


Figure 5.7: Change in regional distribution of load as allowable deviation from load allocation targets is reduced from 100MW to 20MW

Figure 5.7 shows how the distribution of load changes in the next phase of shortfall reduction. Load supplied in Eko DisCo increases but this comes at a disproportionate cost in Enugu region. Eko is in the South West, where there is a concentration of generators, but also a concentration of other DisCos with relatively large loads.

The distribution of load in the cases in the third phase of shortfall reduction are shown in figure 5.8. At this point Yola DisCo is still below its target but any increase in supply to Yola is offset by a disproportionately large reduction in supply to other DisCos. Yola is an outlying DisCo on a radial line without local generation. There is an upper limit on how much load can be supplied in this DisCo. So loads in other DisCos are arbitrarily disconnected until the load supplied in Yola is the correct proportion of the total load supplied.

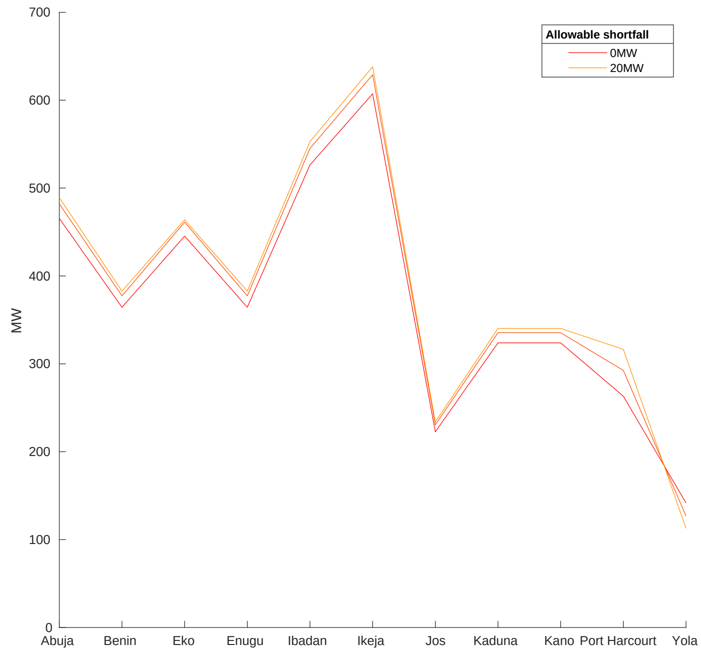


Figure 5.8: Change in regional distribution of load as allowable deviation from load allocation targets is reduced from 20MW to 0MW

This example shows that the deviation from the load allocation targets can be substantially reduced at relatively little cost in terms of extra load shedding. However, there are cases when hard transmission limits mean that adhering to the targets is not possible without substantial reduction in supply to some DisCos and even, in some cases, arbitrability restricting supply so as to meet the constraint in proportional terms without increasing load served in any DisCos. This illustrates the need to enforce the targets in a flexible way.

Short/Long Term Load Allocation

In the previous section we gave an illustrative example of the tension between the load allocation targets and maximizing load served. In this section we quantify this tradeoff for a set of operating points. We then ask if we can do better by enforcing load allocation over longer time scale.

Using the method described in Section 5.2.2, we have obtained the Pareto frontiers of the AC Load Allocation problem for all of the 21 operating points that were selected to approximate the actual operating conditions of the network. An illustrative selection of these are shown in Figure 5.9. The horizontal axis in this chart measures the total deviation from the load allocation targets, while vertical axis measures the total load supplied. The different curves show Pareto frontiers between the two objectives for different generation operating points.

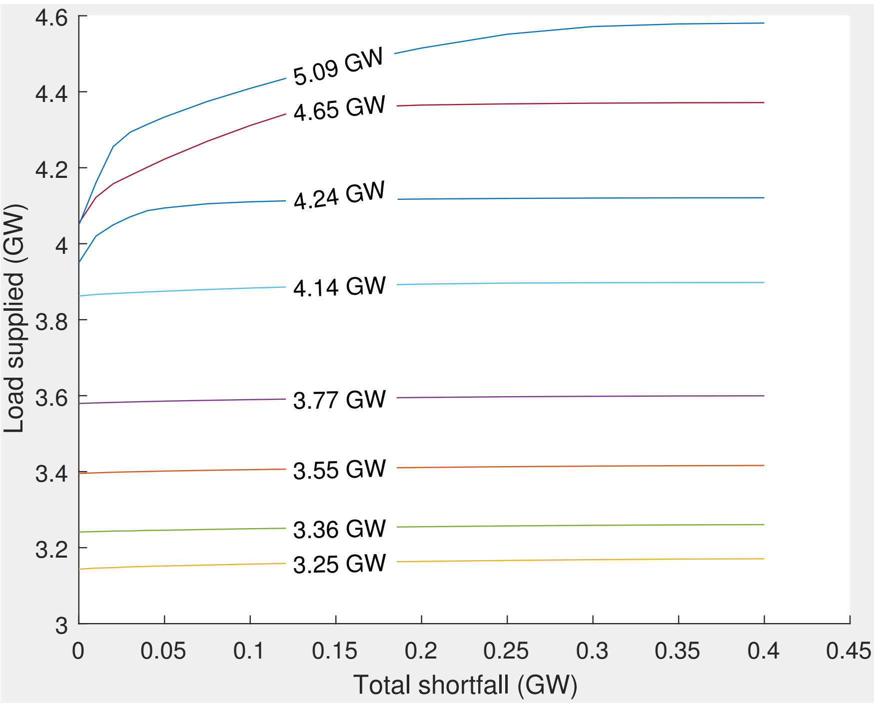


Figure 5.9: Pareto frontiers for an illustrative selection of individual operating points labelled by total available generation

If the generation availability level is very low, the Pareto frontier is nearly flat, which means that shortfall, as defined in constraint (5.20), can be eliminated with little cost in terms of total power supply. In these operating points power can be allocated to where policy makers have decided it should be without causing network congestion. The small decrease in overall power supplied is caused by slightly higher transmission loses. However, for the operating points with more available generation it is not possible to supply enough power to some DisCos, so rigorously adhering to the load allocation targets severely conflicts with maximizing supply. This can be seen from the steep upward sloping curved region at the left-hand end of these curves which indicated that gains in terms of reduced shortfall come at the cost of a significantly reduced overall power supply.

The particularly steep curves at very high generation availability levels illustrates a potential downside of a proportional policy of the kind described in this paper: the potential to create perverse incentives for the SO. If a DisCo is in shortfall relative to its proportional target, but there is a hard transmission constraint preventing further power being supplied here, then the only way for that DisCo's proportional target to be achieved would be to reduce the power supplied to other DisCos. In the Nigerian case we find that this happens in some individual operating points with high generation availability (over 4.5 GW) and a large weight on the objective of minimizing shortfall. Available generation is very rarely this high in Nigeria, but if availability improves without expansion of transmission capacity, proportional regulation of this kind may cause problems.

Ultimately the load allocation targets are incompatible with maximizing supply because some DisCos are unable to absorb enough power. Nevertheless, at the current average level of available capacity, it is possible to trade off load supplied in one DisCo vs another without changing the overall level of supply that much. As available generation increases, the trade-off between regional fairness and total supply is more difficult. Supplying more load in outlying DisCos comes at a substantial opportunity cost in terms of overall supply.

We next compare the Pareto frontier for the Short-term and Long-term AC Load Allocation problems using the methods described in Sections 5.2.2 and 5.2.3. We have obtained 20 solutions to the Short-term Load Allocation Problem. The shortfall averaged over all operating points ranges from 2 MW to 380 MW. Reducing the average shortfall to 2 MW comes at the cost of 5% additional load shedding.

However, average shortfall as defined in the Long-term Load Allocation problem can be reduced to a similar level with much less load shedding. In our solutions to the Long-term problem the lowest shortfall obtained is 5 MW, which comes at the cost of only 2% of additional load shedding (see Table 5.4). Although our solution method for the Long-term Load Allocation problem can in theory leave a duality gap, we can see from the duality gaps reported in Table 5.4 that these points are very close to optimal. As a percentage of the objective function value the largest gap is 0.05% and most are much smaller than this.

In Figure 5.10 we plot all of the solutions obtained to the Short-term and Long-term Load Allocation Problems in terms of the average shortfall in each operating point. In this figure we approximate the Pareto Frontier by interpolating between these points.

For both problems there is a flat region on the right of the Pareto frontier indicating that these levels of shortfall can be reduced at little cost in terms of total power supply. However, in both cases there is also a steep downward sloping curved region at the left-hand end, indicating that it is not possible to adhere to the load allocation targets without reducing the supply to DisCos where more could be supplied.

In the Short-term problem the transition between the flat and sloped regions of the Pareto frontier occurs when the average shortfall has been reduced to 140 MW. This is a higher level than for the long-term problem where the transition occurs around 50 MW. Further reducing shortfall below these levels comes at an increasing cost in terms of load shedding. This implies that enforcing load allocation targets over a longer time horizon is substantially more efficient, as may be expected. However, this also means that some regions will be under supplied for a longer period, which may not be acceptable.

Table 5.4: Solutions to the Long-term Load Allocation Problem

W	0.001	0.01	0.05	0.1	0.3	0.5
Supply (GW)	4.040	4.039	4.036	4.030	4.020	4.012
Supply (% of maximum)	100%	100%	99.9%	99.7%	99.5%	99.3%
Shortfall (GW)	0.431	.334	0.194	0.104	0.048	0.026
Duality Gap (%)	0.00%	0.00%	0.00%	0.01%	0.03%	0.02%

W	0.8	1	2	4	6	6.5
Supply (GW)	4.009	4.008	4.003	3.993	3.969	3.958
% of maximum	99.2 %	99.2%	99.1%	98.8%	98.3%	98.0%
Shortfall (GW)	0.022	0.019	0.015	0.012	0.007	0.005
Duality Gap (%)	0.05%	0.01%	0.01%	0.01%	0.01%	0.01%

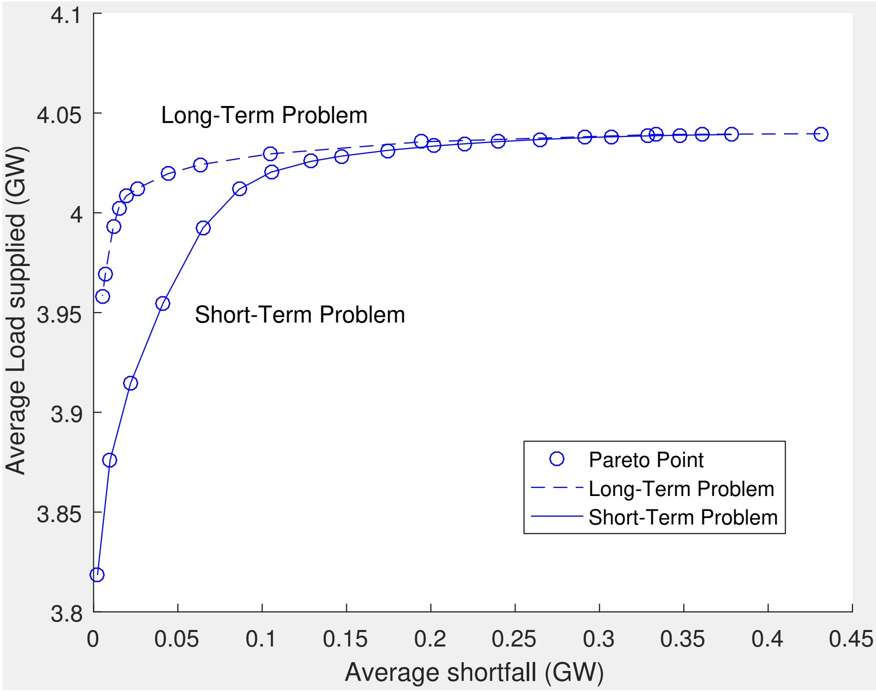


Figure 5.10: Pareto optimal points of the Short-term and Long-term Load Allocation Problems

Sensitivity to Estimate of Unsuppressed Load

The level of unsuppressed load (demand without load shedding) is the parameter of the model that has the greatest range of uncertainty. Therefore, sensitivity analysis of these results was carried out to quantify the sensitivity of our model to potential inaccuracy of this estimate. We might expect that despite the transmission constraints, some regions of the network can absorb more power. If so, the total load that can be supplied may also increase if the loads in these regions are increased.

The OPF model was solved for 21 individual operating points with overall load level ranging from 5 GW to 15 GW increments of .5 GW (Note, these are not the same 21 operating points used in the case study). In the first instance the optimization was carried out with an objective to minimize load shedding with no load allocation targets. In the second case we require strict adherence to the proportional load allocation targets. The individual loads at each bus were scaled uniformly. The analysis was carried out for 6 levels of generation availability: 3, 3.5, 4, 4.5, 5 and 5.5 GW. In each case we scale the nameplate real and reactive power generation capacity of each generator by the same factor so that the spatial distribution of generation is the same.

The results of maximizing total supply are shown in figure 5.11 where total load supplied is plotted against unsuppressed load for all 6 scenarios. A horizontal line for a given scenario means that the level of load supplied does not increase if the level of unsuppressed load is higher. The 5 and 5.5 GW generation availability operating points display the highest sensitivity to the load level. The 3, 3.5, 4 GW operating points display very little sensitivity to the level of load while the 4.5 GW scenario displays some sensitivity to the level of load between 5 and 7 GW.

The results when we minimize the shortfall from the proportional targets are shown in figure 5.12. Again, the sensitivity to load level is largest in the lower range of unsuppressed load (5 to 10GW). Note that with strict adherence to load allocation targets it was not possible to find any feasible solutions for the lowest generation scenario with 3 GW of available capacity.

As shown in Figure 5.2 generation availability in the study period is very rarely in excess of 4.5 GW and never exceeds 5 GW. Therefore, in in the most realistic range of generation scenarios (3 to 4.5 GW) our results are not sensitive to the estimate of load if it is in excess of 7 GW. Based on discussions with consultants working in the sector we have verified that demand is always likely to be in excess of 7 GW. We conclude that our results would not be significantly affected by considering a range of different demand estimates or variation in demand across the day.

However, the methods we have developed could easily accommodate a situation in which modelling variation in demand over the day is more important. For example, one could replicate each generation operating point with three operating points with a peak, off-peak and medium demand estimate and weight these in the objective function according to the proportion of the day they represent. The methods of Section 3.5 could be used to obtain these demand estimates. In order to do this, we would need estimates of the number of customers, split by tariff category at each grid supply point, or at least in each distribution company area. In principle we could use these in combination with the coefficients of the DSR model we estimated in Chapter 3 to estimate the total unsuppressed demand in Nigeria. However, these coefficients are probably not representative of the whole of Nigeria because they were estimated using data from only one DisCo. For this reason, and the lack of customer numbers at grid supply points we did not attempt this method of demand estimation in this thesis.

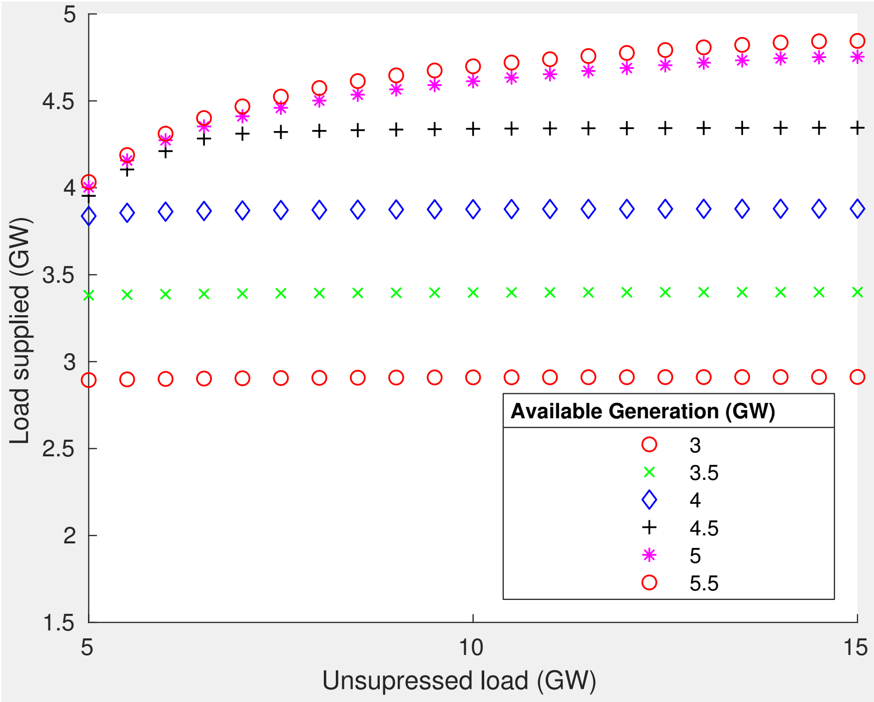


Figure 5.11: Total load supplied as a function of unsuppressed load without load allocation targets

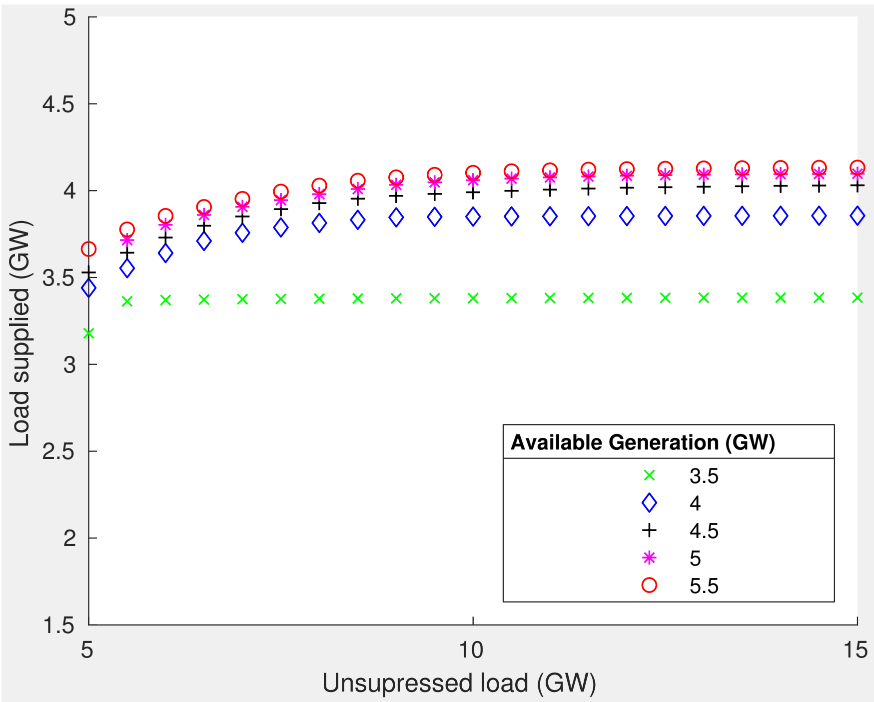


Figure 5.12: Total load supplied as a function of unsuppressed load with load allocation targets strictly enforced

Figure 5.13: Proportion of network components on limits plotted against available generation capacity for 397 generation scenarios in study period

Transmission Network Constraints

We have shown that the transmission capability of the system is about 4500 MW when we try to maximize load delivery. We now consider what factors constrain transmission at this level. For every generation scenario in Figure 5.2 we solved the single period OPF to minimize the load shed without considering the load allocation targets. We calculate the proportion of which constraints of various types that are at their limits in the solution. We consider real and reactive generator limits, bus voltage limits (split by transmission and distribution level buses) and branch power flow limits (split into transformers and transmission lines). Figure 5.13 shows the proportion of each type of constraints on their limits plotted against the available capacity in that scenario.

The binding constraint does not seem to be the capacity of transmission lines, since very few transmission lines are at capacity limits, even for scenarios with high available generation. At a critical level of available generation at about 4200MW, the proportion of transmission voltage buses on their voltage limits rapidly rises. Below this level of available capacity, we note that the proportion of distribution level buses on voltage limits and transformers on capacity limits increases in proportion to the level of available capacity. For low levels of available capacity, as certain local regions of the system become congested, it is possible to supply more power to outlying areas to increase the load supplied. But for high levels of available capacity the variations in voltage magnitude required to supply more power to outlying areas become too high.

The concentration of generation assets in the south of the system creates the need to transmit both real and reactive power over long distances. It seems that there is enough capacity for the power flow at the current levels of generation. However, the need to transmit large amounts of reactive power creates a large voltage drop. Providing some source of reactive power, such as capacitors, in those parts of the network could help to solve this problem.

5.4 Conclusion and Policy Implications

In a power system in a long-term power shortage, some policy must be adopted that describes how to ration available power. If the SO is only incentivised to maximize delivered load, say by payments for transmission services, then they are unlikely to distribute power equally to different parts of the system, which may be unfair or otherwise politically unacceptable. Countries with such power systems, such as Nigeria, therefore, specify load allocation targets or similar metrics for each region. SOs are then penalized for failing to meet these targets, incentivizing them to choose a more equal distribution of the available power.

However, as we have argued in this paper, this can come at a substantial cost in terms of additional load shedding. Depending on the design of the policy the SO may be incentivized to minimise shortfall at every operating point or attempt to balance shortfall or surplus over the long term. It may also be allowed to deviate from load allocation targets to a certain extent. Policy makers need to know what the efficiency penalties are in each of these cases to design an optimal load allocation policy.

We have developed methods which can be used to quantify the trade-off between different objectives and implement a chosen policy and applied this to the Nigerian power system. Our case study shows that the cost of the Nigerian load allocation targets is substantial, as it reduces the total amount of power supplied by up to 5%. This does not imply that the policy is inefficient, if a more equal distribution of the available power has substantial benefits, but without this quantification a trade-off between fairness and efficiency cannot be made.

It could be argued that instead of finding a balance between proportionality and maximum load, the SO should attempt to maximize a utility function of power supply. The utility of power in each region of the network is likely to have diminishing returns, so maximizing the sum of the utility functions in all regions is likely to enforce some level of proportionality between regions in any case.

Alternatively, the SO could have fixed targets for the minimum level of power to be supplied in each region, rather than proportional targets such that the target moves, in absolute terms, with the total level of power supply. Once these targets have been met, surplus power could be distributed according to some other rule. This could solve the problem where proportional regulation encourages the SO meet the load allocation target in a given region by shedding load elsewhere even though it is not possible to supply more load in the region in question.

Nevertheless, in both these cases there remains a similar difficulty of trying to achieve these objectives subject to different network conditions as the level and spatial distribution of available generation varies. Furthermore, both objectives have short-term and long-term versions analogous to the Short-term/Long-term Load Allocation Problems described above. Therefore, the methods we describe in this paper could be adapted to these alternative problems.

In the course of our investigation we found that investment in reactive compensation could make it feasible to achieve the Nigerian load allocation targets with less load shedding. However, it should be noted that although reactive support is cheap relative to generation or transmission lines, developing world power systems often find it difficult to meet even this level of investment. It is therefore important to consider how to most efficiently operate and regulate the existing network. Although alleviating these problems by investment in reactive support is beyond the scope of this paper our method could be used to evaluate different plans for investment in reactive support. An interesting subject of further research would be to optimize the location of capacitor banks.

Chapter 6

Conclusions

In this thesis we analysed the difficulties faced by various actors in power systems during a chronic power shortage. A long-term solution to this problem requires new investment, however the persistence of these problems in many parts of sub-Saharan Africa and South Asia means that research on shorter-term operational aspect of this problem is also needed. This is an area that has been sorely neglected in the power system modelling literature. We analysed three areas: Demand forecasting, national level power rationing and distribution level demand management. We now review our conclusions in each of these areas, discuss how these results could be applied in practice and suggest where further research should focus.

In summary this thesis has made a number of contributions that will be of interest in the context of power systems operating in a chronic power shortage. Firstly, we developed accurate short-term forecasting methods for electricity demand in chronic power shortage conditions. These methods also have applications for post-hoc reconstruction of the unsuppressed electricity demand. Secondly, we formulated a decision support model to help DSOs plan load shedding using the multi period stochastic knapsack problem. Thirdly we developed methods that the SO or regulator can use in a power shortage to quantify the trade-off between maximizing the total amount of power delivered and distributing the available power in a fairer way. We quantified the trade-off between regional equity and total power supply in the specific case of Nigeria, showing that current Nigerian policies reduce the total amount of power delivered by up to 5%.

We have also made some general methodological contributions which will be of interest outside the field of energy system modelling. We showed how to extend time varying coefficient regression models to the case of periodic variation in the coefficients modelled by trigonometric terms. We extended the stochastic knapsack problem with random weights to a multi-period case where the random weights are the output of a stochastic process and the forecast variance is endogenous to the model. Although this problem is difficult to solve in general, we gave a tractable formulation in the case where the stochastic variation in the item weights are the output of an autoregressive process.

The models described in Chapter 3 clearly show that it is possible to provide accurate forecasts of the suppressed demand conditional upon a particular policy of load shedding. We developed two methods, which we call the ASM and DSR approaches. We noted that the underlying model assumptions of the LGSSM seem to be poorly satisfied for the ASM model. Consequently, the probability distributions derived from the model are flawed. Despite this, the ASM models are able to provide more accurate point forecasts than the DSR model. For many of the feeders we modelled there is evidence of non-stationary and non-Gaussian standardised forecast errors. In Section 3.6 we described a program of further research to determine if these violations are inherent to the data generating process or whether these are artefacts of poor data quality and lack of the required data to account for structural breaks. The DSR model could potentially be improved by incorporating additional data such as socio-economic characteristics of customers or estimates of the number of illegal connections. Utilities that want to use this approach should therefore focus on obtaining accurate and detailed information about the population supplied by each feeder.

The DSR model also has the advantage that it can be applied to feeders where demand data has not been recorded. This makes this model particularly attractive to estimate the unsuppressed demand at a national level. Existing methods for this rely on unreliable proxy variables or bottom up up calculations that are difficult to verify. In Nigeria TCN goes as far as to periodically carry out "stress tests" to establish the maximum feasible supply to each DisCo without generation constraints. This involves maximising supply to a single DisCo by enforcing abnormally high levels of load shedding in all other DisCo. This provides some empirical measure of the unsuppressed load but, as well as being highly disruptive it is complicated by transmission and distribution constraints which most likely lead to underestimation of demand.

The need for "stress testing" of the system could be removed by combining the methods of Chapters 3 and 5. An estimate of the unsuppressed demand for the whole country produced using a dynamic regression model could be incorporated into an up to date transmission model. This would provide the data for applying our methods to find the maximum supply possible to each DisCo.

It would also be of great interest to apply our demand estimation methods to a dataset coving a longer time period. If a sufficiently long-term load dataset could be obtained (i.e. years) then it would be possible to apply econometric methods to the resulting estimates of unsuppressed demand. This would enable medium-term predictions of demand growth to be made.

Further exploration of this application of dynamic regression modelling should first explore in more detail whether our DSR model is the best way of detecting the relationships between the number of connected customers, the observed load and the latent demand. An alternative way of forming a DR model with periodicity in the coefficients is the dynamic factor model of [15]. To compare these models, we suggest that a study should be carried out to see which model can best recover the details of a simulated unsuppressed demand time series from the suppressed demand.

It is interesting to consider how probabilistic forecasts of suppressed demand could be incorporated into the operations of DSOs. We argued in Chapter 4 that the relevant considerations for a DSO can be incorporated into an optimisation framework that resembles a stochastic knapsack problem. This framework allows DSOs to prioritise feeders while also managing risk of overconsumption. The existing literature has focused on finding exact solution methods to the stochastic knapsack problem. However, since software packages that implement these methods are not available it is useful to formulate an approach that can be easily implemented and solved using standard solvers. To this end we described a MILP approximation to stochastic knapsack problem. We propose an extension to the stochastic knapsack problem where the demand of the feeders evolve according to a stochastic process. This allows us to model the repeated character of the DSO's problem.

Part of our motivation for exploring this problem is the potential theoretical interest of a class of problem where the probability distributions are affected by factors that are endogenous to the model. However, at least in the case we explored (autoregressive processes with Gaussian noise) modelling the effects of decisions on future problems seems to be of limited importance to obtain a good policy. Further analysis is required to see how this affected if the item weights (i.e. demand) are determined by different stochastic processes or probability distributions.

Progress in this area would be enabled by assembling billing and revenue collection data at a feeder level. This should be linked with demand and load shedding data (of the type we used in this thesis) as well as records of the SO's instructions. This would allow bench-marking of the current performance of DisCos with respect to demand management. Having collected this data we suggest that further research should focus on the question: What level of demand can the DisCo "commit" to connecting in advance on the basis of long term forecasts of demand and supply such that they can avoid disconnecting "committed" feeders with a given probability? This requires the DSO to plan based on long term forecasts. This would be useful because it potentially would allow DSOs to offer a more predictable (but still intermittent) supply to customers.

We argued in Chapter 5 that the existence of transmission constraints in addition to limited generation causes conflict between the objective of minimising load shedding and distributing the burden of load shedding in a (geographically) equitable way. If the SO is only incentivised to maximise delivered load, say by payments for transmission services, then they are unlikely to distribute power equally to different parts of the system, which may be unfair or otherwise politically unacceptable. Countries with such power systems, such as Nigeria, therefore, specify load allocation targets or similar metrics for each region. However, this can come at a substantial cost in terms of additional load shedding. Depending on the design of the policy the SO may be incentivised to minimise shortfall at every operating point or attempt to balance shortfall or surplus over the long term. It may also be allowed to deviate from load allocation targets to a certain extent. The methods developed in this thesis can be used to quantify the trade-off between different objectives and implement a chosen policy. Our case study shows that the cost of the Nigerian load allocation targets is substantial, as it reduces the total amount of power supplied by up to 5%.

Chapter 7

Summary of Sustainable Energy For All Rapid Assessment and Gap Analyses

Table 7.1: Summary of Sustainable Energy For All Rapid Assesment and Gap Analyses

Country	Electricity Access	Power quality	Non-technical losses
Angola	30% electricity access, overwhelmingly in capital, Luanda. Diesel widely used to supplement power sector. Diesel price regulated and subsidized.	Infrastructure has not kept pace with consumption, leading to power rationing.	Tariffs cover less than 20% of the electricity costs. Losses above 40%.
Gambia	Unaffordable electricity, high tariff relative to other countries due to reliance on diesel, poor fuel supply to generators.	Unreliable grid electricity	Distribution losses caused by inefficiencies in main utility.

Country	Electricity Access	Power quality	Non-technical losses
Ghana	72% Electricity Access	Load shedding has periodically been caused by unreliable fuel supply to thermal generators and low rainfall	Distribution losses as a result of weak regulatory and enforcement capacity. Policy in early 2000s to move to cost reflective tariff.
Kenya	Interconnected grid and 14 minigrids. 50% urban electrification, 6% rural (30:70 population split). 50% connection subsidy, first 50kWh/month subsidized. Large contribution hydro and geothermal (55%) + much more potential.	Supply<Demand over last decade. Substantial growth in demand due to economic growth.	Reform and restructuring since mid 90s. Total losses at 17.3% in 2012.
Lesotho	26% national grid electrification (65% urban, 6% rural).	Peak demand 2 * available capacity, largely due to increased mining.	Tariffs not cost reflective.
Liberia	Very low electricity access (1%). Liberia Electricity Company operates a small grid in the capital district.		
Namibia	70% urban electrification, 25% rural (30:70 population split)	Peak demand twice installed capacity, imports remainder from SA. Recently SA has its own power shortage.	7% losses.

Country	Electricity Access	Power quality	Non-technical losses
Nigeria	47% population access to grid electricity.	Experiencing power crisis. 45% installed capacity out of service.	Low tariffs, not cost reflective. High incidence of power theft and non-payment of bills.
Rwanda	20% population grid connected	Some load shedding in peak hours.	
Sierra Leone	10% electricity access. Isolated electricity supply systems, mostly destroyed in war. Electricity consumption mostly in Freetown.	Frequent blackouts.	
South Sudan	1% population access grid electricity. 3 isolated grids in population centers critically short of fuel and spare parts. Most electricity self generated.		
Swaziland	61% access. Local hydro generation + imports from SA supply grid.	Shortages caused by insufficient installed capacity as well as insufficient dam capacity and unreliable rainfall	14% distribution losses.
Tanzania	18.4% population access to grid electricity.	Demand increasing at twice rate of generation expansion. Severe power rationing and blackouts	Tariff does not cover costs. Problems with revenue collection. 23% losses.

Country	Electricity Access	Power quality	Non-technical losses
Zimbabwe	38% electricity access 83% urban 13% rural	Power shortage. 35% capacity out of service. Requires imports from Mozambique but affected by wider SAPP power shortages.	Non cost reflective tariff. Inefficient electricity billing and collection
Bangladesh	47% access	Load shedding in peak hours	“Endemic” power theft
Bhutan	82.5% access via grid and off grid solutions	High reliability	
Cambodia	35% access. Limited national grid and isolated Rural Electricity Enterprises	Load shedding and unplanned blackouts are common	Non cost reflective tariff on national utility. REEs tend to recover costs but are highly expensive for consumers and losses are higher.
Fiji	82%. Several isolated grids and many minigrids	Generally good availability and power quality	
Malaysia	Nearly universal grid access	Stable and consistent supply	
Mongolia	Almost universal in urban areas, 54% in rural areas		
Nepal	50-70%	Routine load shedding, worst in dry season when run-of-river hydro limited	
Pakistan	60% grid connected	Chronic load shedding due to insufficient installed capacity	Problems with power theft and non payment of bills

Country	Electricity Access	Power quality	Non-technical losses
Sri Lanka	91% households grid connected,	Frequent interruptions due to system faults	

Chapter 8

Experience of Restructuring in South America and South Asia

Chile was the first South American country to pursue reform in 1982. It was followed by Argentina (1992), Ecuador, Peru (1993) and Bolivia (1994), which adopted a similar model of regulation [6]. Other countries also reformed their power sectors including Columbia (1994) and the Central American countries of Panama, El Salvador, Guatemala, Nicaragua, Costa Rica and Honduras (1997) [44]. Diverse countries but all then with low electricity consumption relative to industrialized world.

For the most part electricity sector restructuring in South America has relied on cost based, rather than bid based processes to set prices, with an ISO dispatching generation according to marginal cost. The primary exception in Columbia which introduced a bid based spot market in 1994. The Chilean/Argentine model introduced wholesale competition

Of primary interest to Sub Saharan African countries will be the measures by which the many countries in South America have substantially reduced non-technical losses. Throughout South America completion was not considered feasible at the retail level, so reform focussed on performance based regulation for monopoly retail and distribution companies [44]. These companies have in some instances been able to deliver remarkable reductions in non-technical losses.

In a report [57] the World Bank singled out some particular cases.

8.0.1 DELSUR in El Salvador

In late 90s 5 major Distribution Companies in El Salvador privatized. Delsur was privatized 1998 and Purchased by the U.S. group PPL Global Inc. The new owners improved management information systems and customer databases to coordinate and field campaigns to improve metering, billing and collection. Commercial functions were made more transparent and efficient. Losses, at 15% in 1998, were reduced to 7% by 2002.

8.0.2 Enersis of Chile

Chilectra Metropoklitana was privatised 83-87 in early electricity restructuring. Its owners, Enersis have since participated in privatisation of utilities in Argentina, Chile, and Columbia, Peru.

Their strategic model focuses on customer segmentation and geographic vectorization of the served area. This allowed them to deliver a 10% reduction in losses in Chile over 7 years, a 14% reduction in Argentina over 4 years, a 12% reduction in Columbia in 3 years and a 7% reduction in Peru in 5 years.

The approach used by Enersis is exemplified by the case of CODENSA in Bogota, Columbia. It should be noted that the punitive actions were only possible because of the cooperation of law enforcement agencies and investment in metering and improved management relied on a tariff structure that allowed an efficient company to cover their costs.

Improved Commercial Management - More points of service to communicate with customers - Collect old debts, allow illicit customers to become normal customers - Improved, metering, billing and collection - Improve external contractor's ethical behaviour and personal safety - Community engagement to improve customer service and promote electricity as a good that needs to be efficiently and commercially managed - Improved MIS and customer database and regularly reconciling it with reality - Allow illicit users to become legitimate customers Technical actions - Tamper proofing infrastructure and meters - Automatic meter reading and monitoring of large consumers - Renovate distribution networks and individual connections of customer connections located in areas with high theft - Installation of meters and replacement of tampered meters Punitive actions - Systematically using legal proceedings against large consumers - Police actions when required - Recovery of old debts using legal system when required - Public information on main cases of electricity theft, in general involving well-known social agents, to promote social condemnation Improved Regulation - Tariff structure to allow efficient DISCOs to recover costs - Sufficient funds generated to support cross subsidisation of consumer classes

8.0.3 Andhra Pradesh State, India

An intensive loss reduction programme combined with restructuring reduced transmission and distribution losses from about 38 percent in 1999 to 26 percent in 2003.

Andhra Pradesh State Electricity Board (APSEB) suffered from large financial losses with inefficiency and theft camouflaged by unverifiable estimates of sales and losses. APSEB was disaggregated into 1 generation, 2 transmission companies and 4 retail and distribution companies. After restructuring the distribution companies embarked on a sustained loss reduction campaign, supported by regulatory and legal changes.

State laws were amended to make electricity theft illegal and rigorous enforcement was backed up by police action where necessary. Police stations provided public notification of all cases of theft. Anti-corruption departments in DISCOs were strengthened. The theft control program focusing on large consumers including tamper-proof meters and protective boxes installed on transformers.

Enforcement activities were backed up by communication and prevention measures. A communication programme informed people about laws, explain utilities' financial situation and effect the of electricity theft on costs and tariffs. Installation of high quality meters, remote meter reading equipment, and higher accuracy meters for large/medium customers and recalibration of existing meters. Furthermore, new Management Information Systems were introduced using centralized customer database to analyse metering, billing and collection performance.

8.0.4 North Delhi Power Limited, India

A public/private partnership (51:49 split between Tata and Government of Delhi) reduced losses from 53% at takeover in July 2002 to 15% in April 2009. Performance-based multi-year tariff regulation allows the company allowed to keep surplus revenue from exceeding performance targets set every 4 years.

NDPL carried out aggressive enforcement activities with "scientific inputs and analysis and a communication programme about dangers of direct tapping from live wires. Newer, higher accuracy electronic meters were introduced and medium voltage distribution networks were introduced in high theft areas. However, 90% of results from advanced metering infrastructure for large customers.

Chapter 9

Fitted models and diagnostic plots for structural models

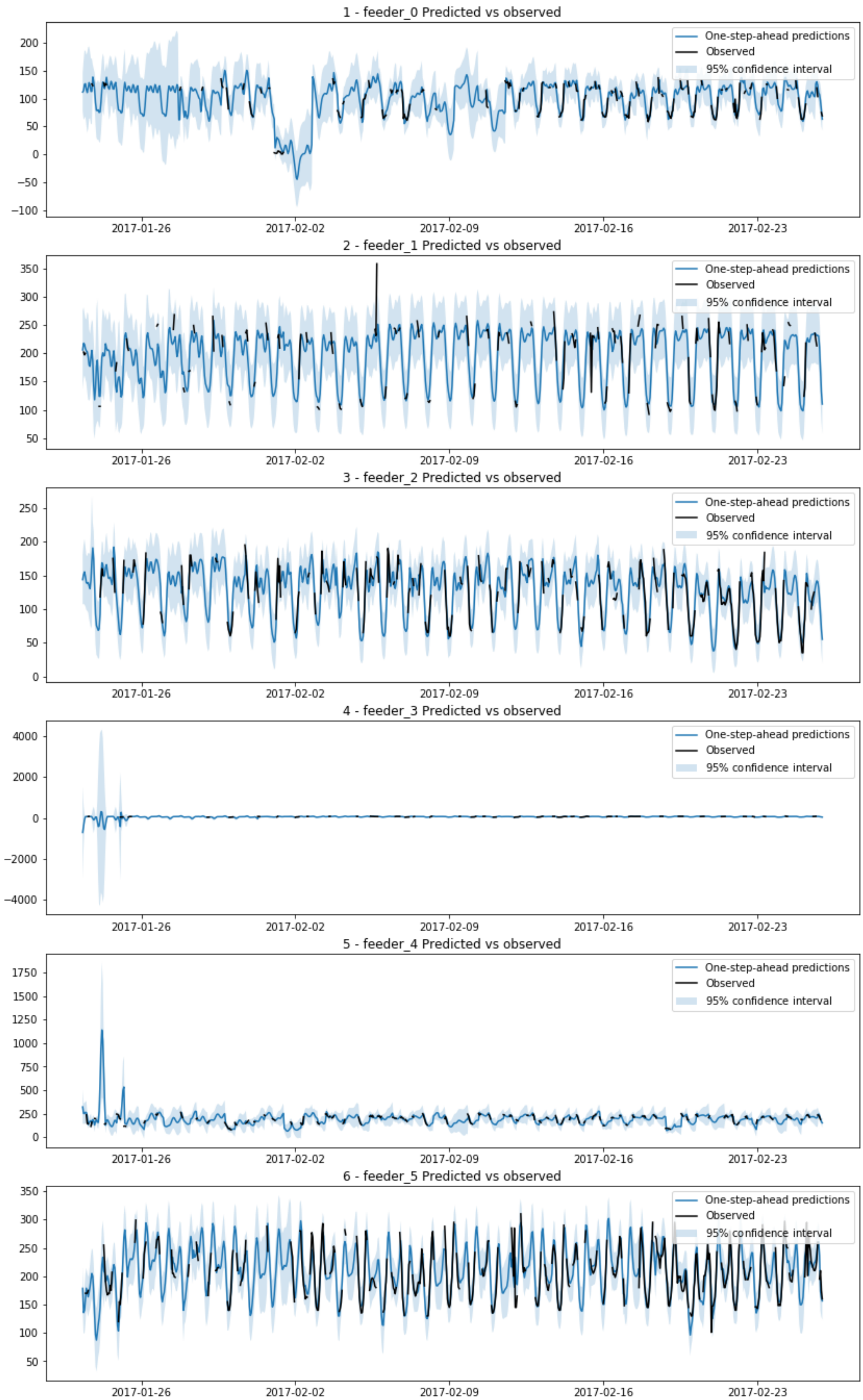


Figure 9.1: Observations compared to fitted values from structural models 0-5

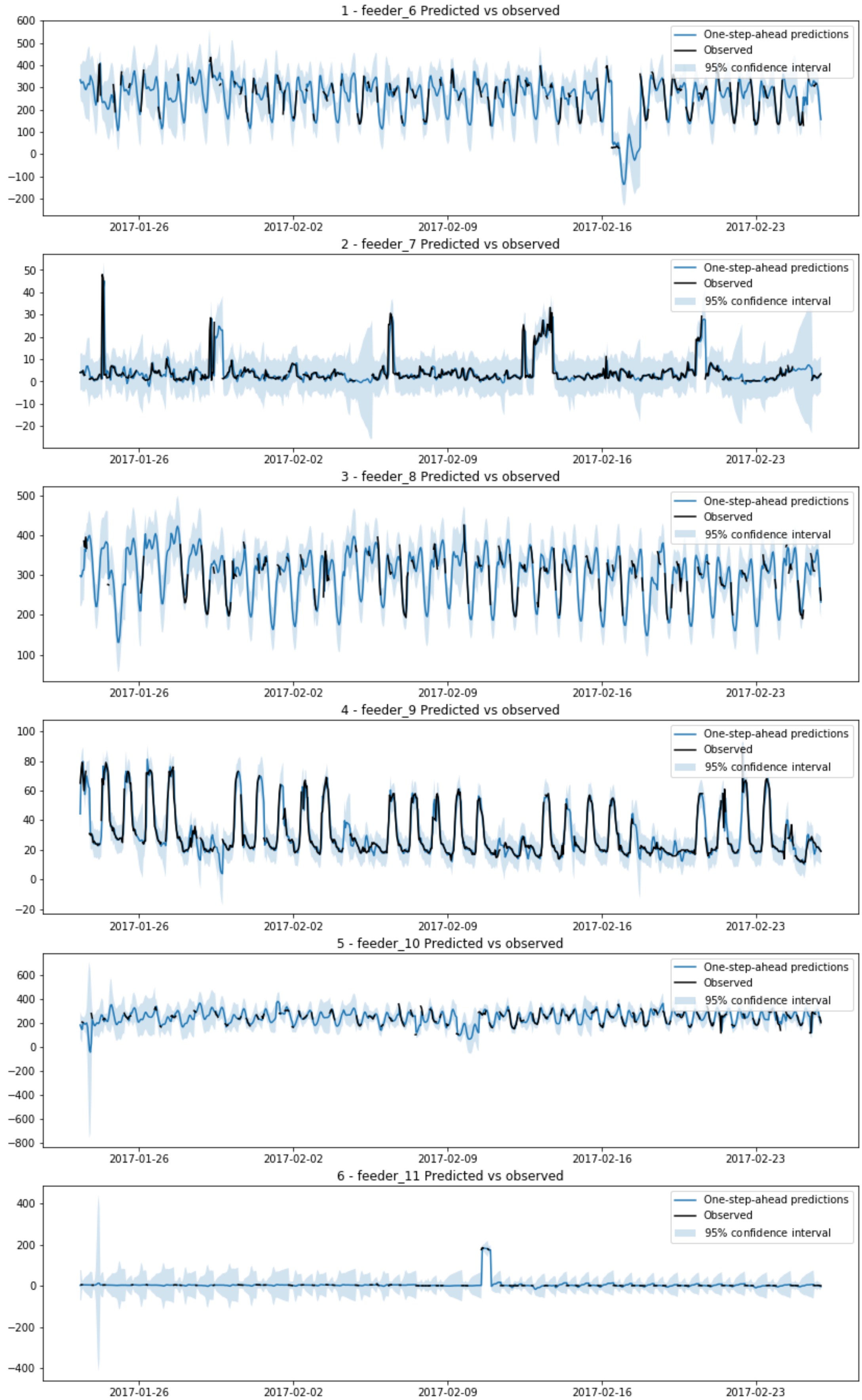


Figure 9.2: Observations compared to fitted values from structural models 6-11

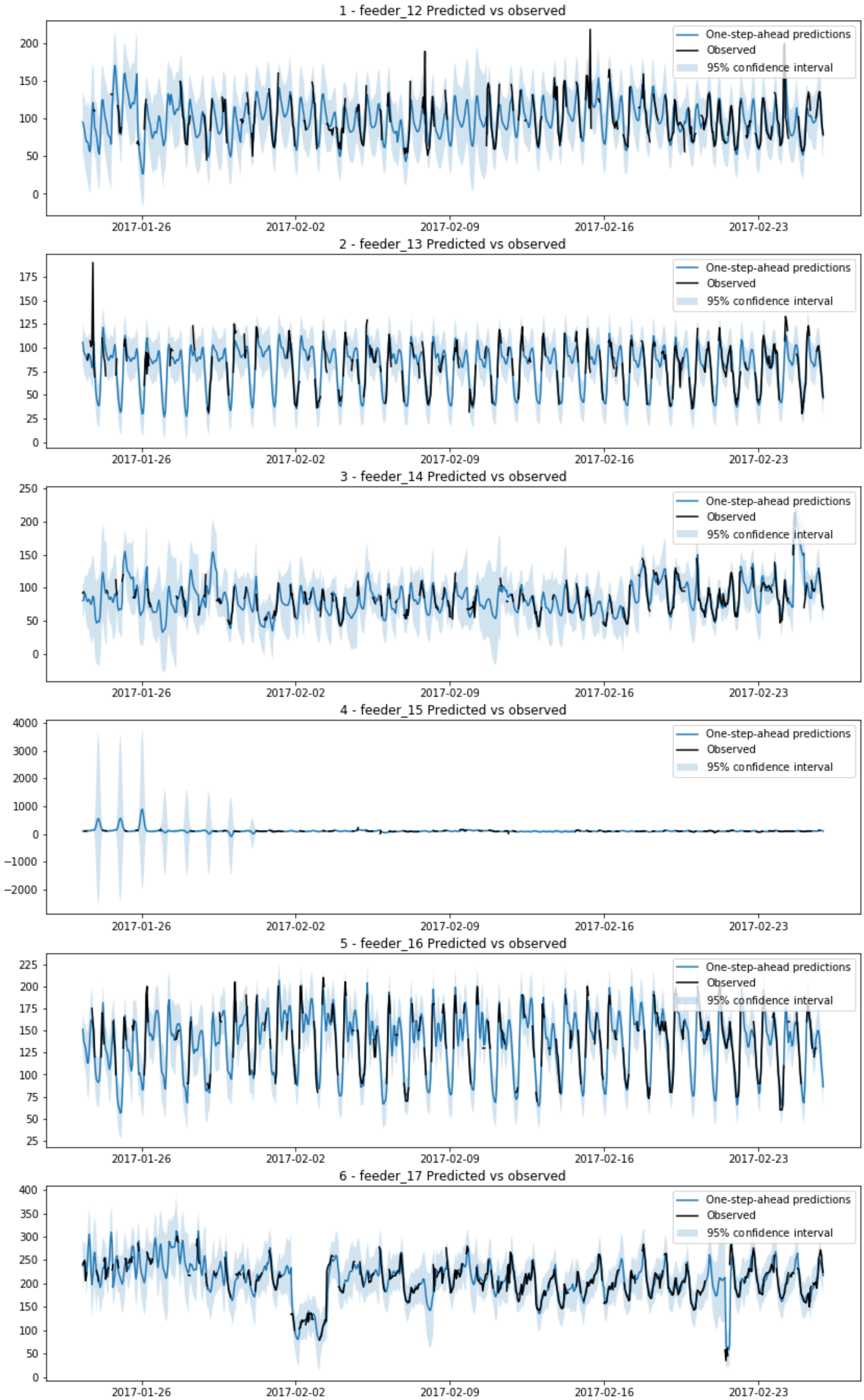


Figure 9.3: Observations compared to fitted values from structural models 12-17

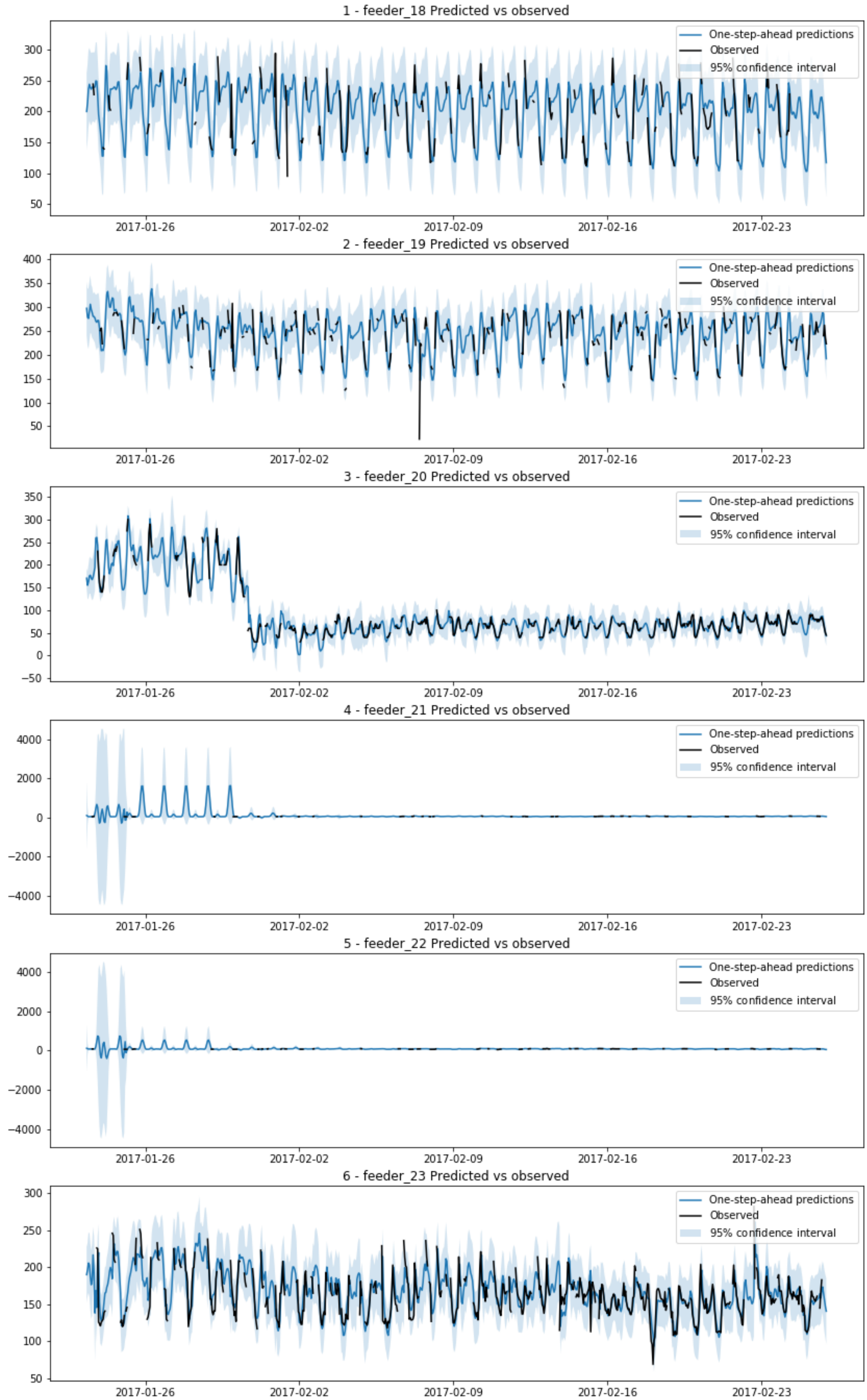


Figure 9.4: Observations compared to fitted values from structural models 18-23

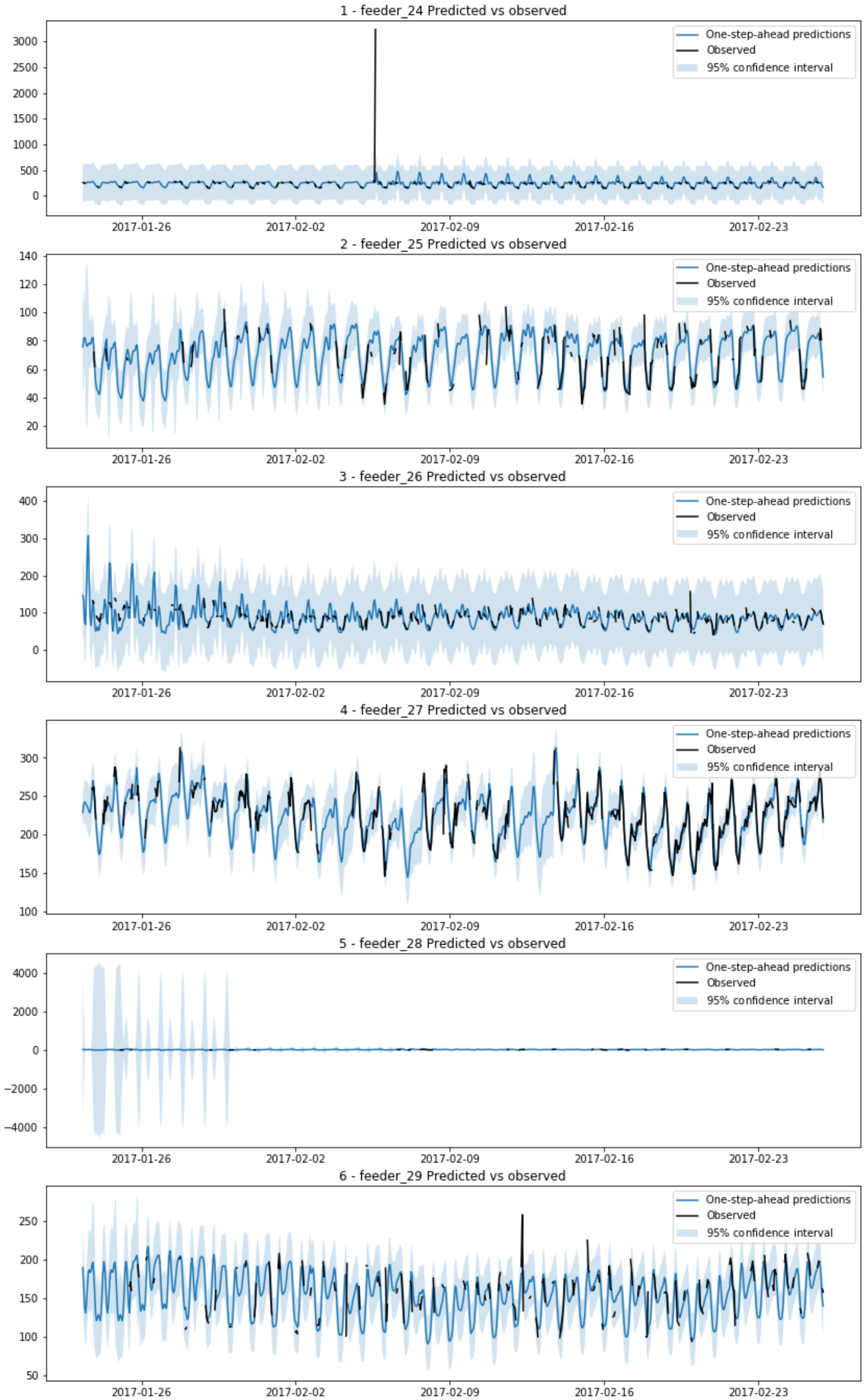


Figure 9.5: Observations compared to fitted values from structural models 24-29

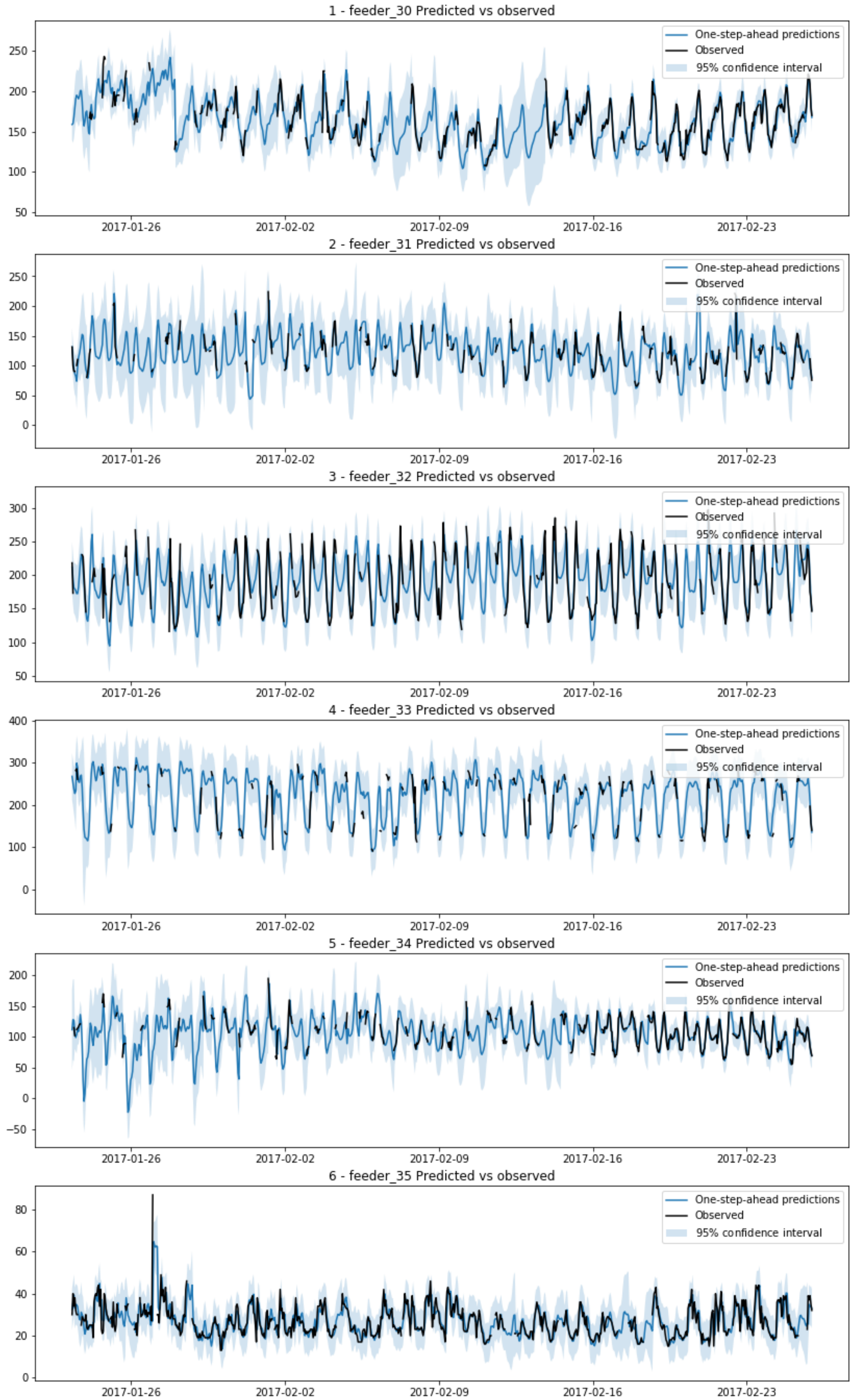


Figure 9.6: Observations compared to fitted values from structural models 30-35

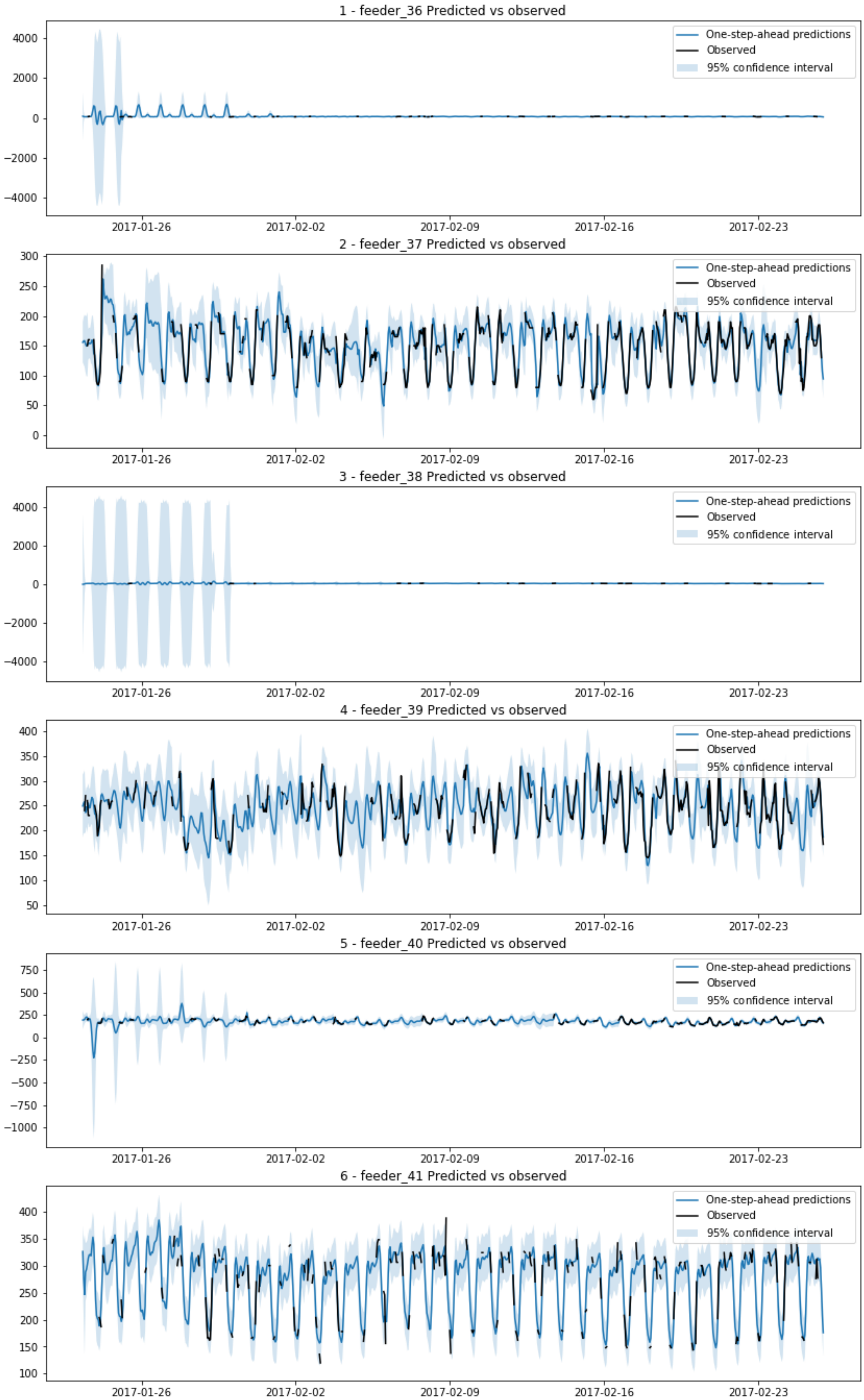


Figure 9.7: Observations compared to fitted values from structural models 36-41

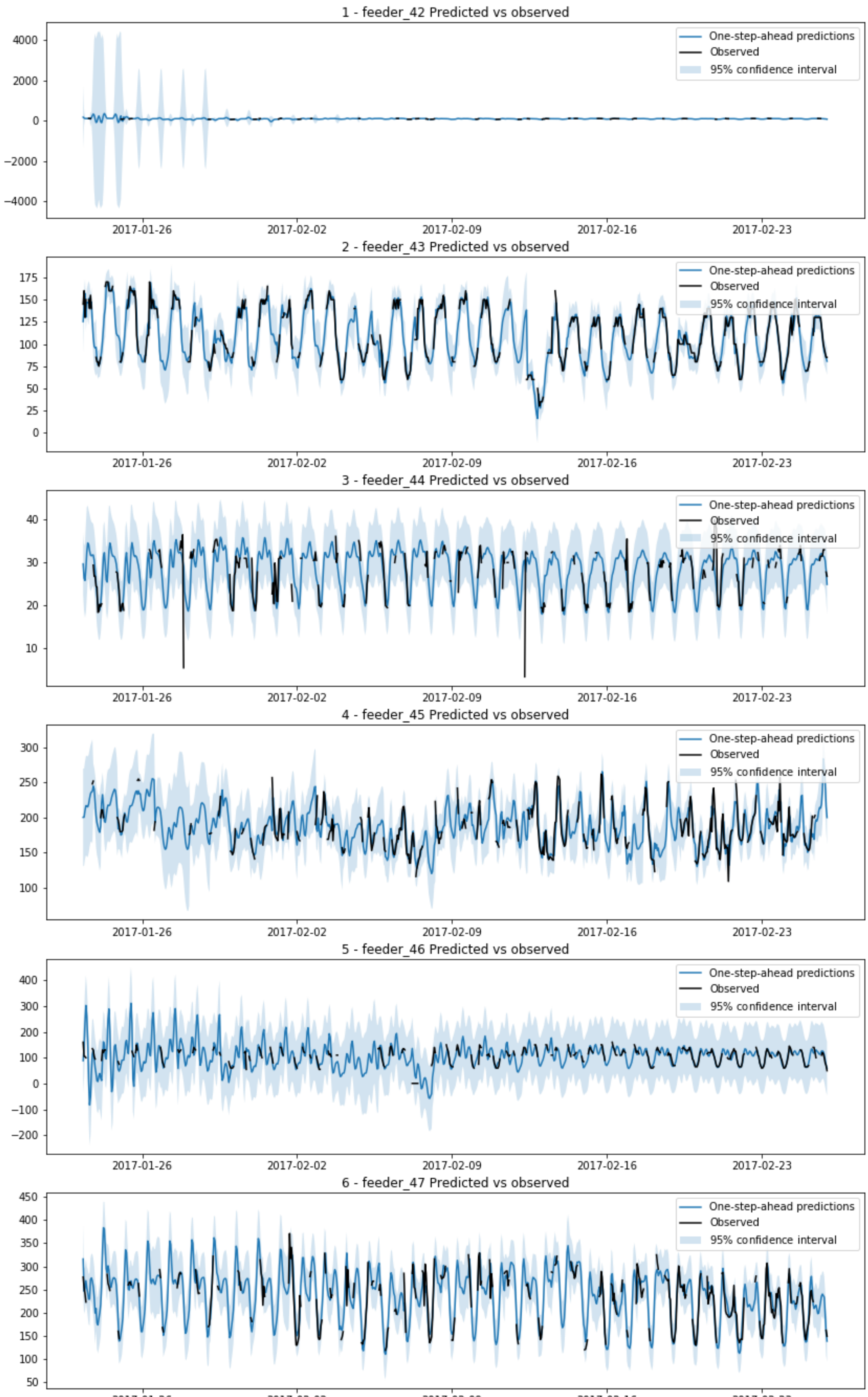


Figure 9.8: Observations compared to fitted values from structural models 42-47

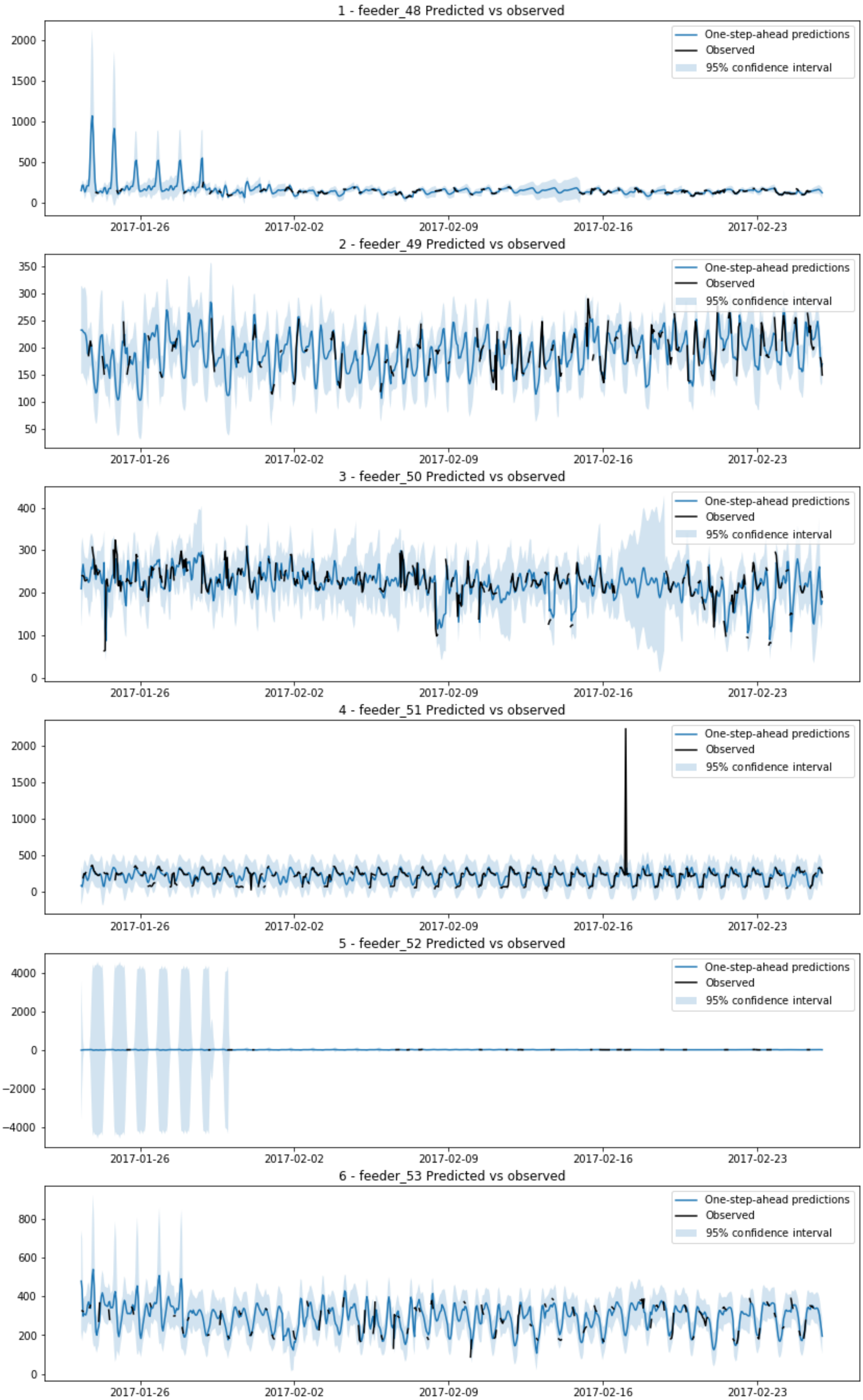


Figure 9.9: Observations compared to fitted values from structural models 48-53

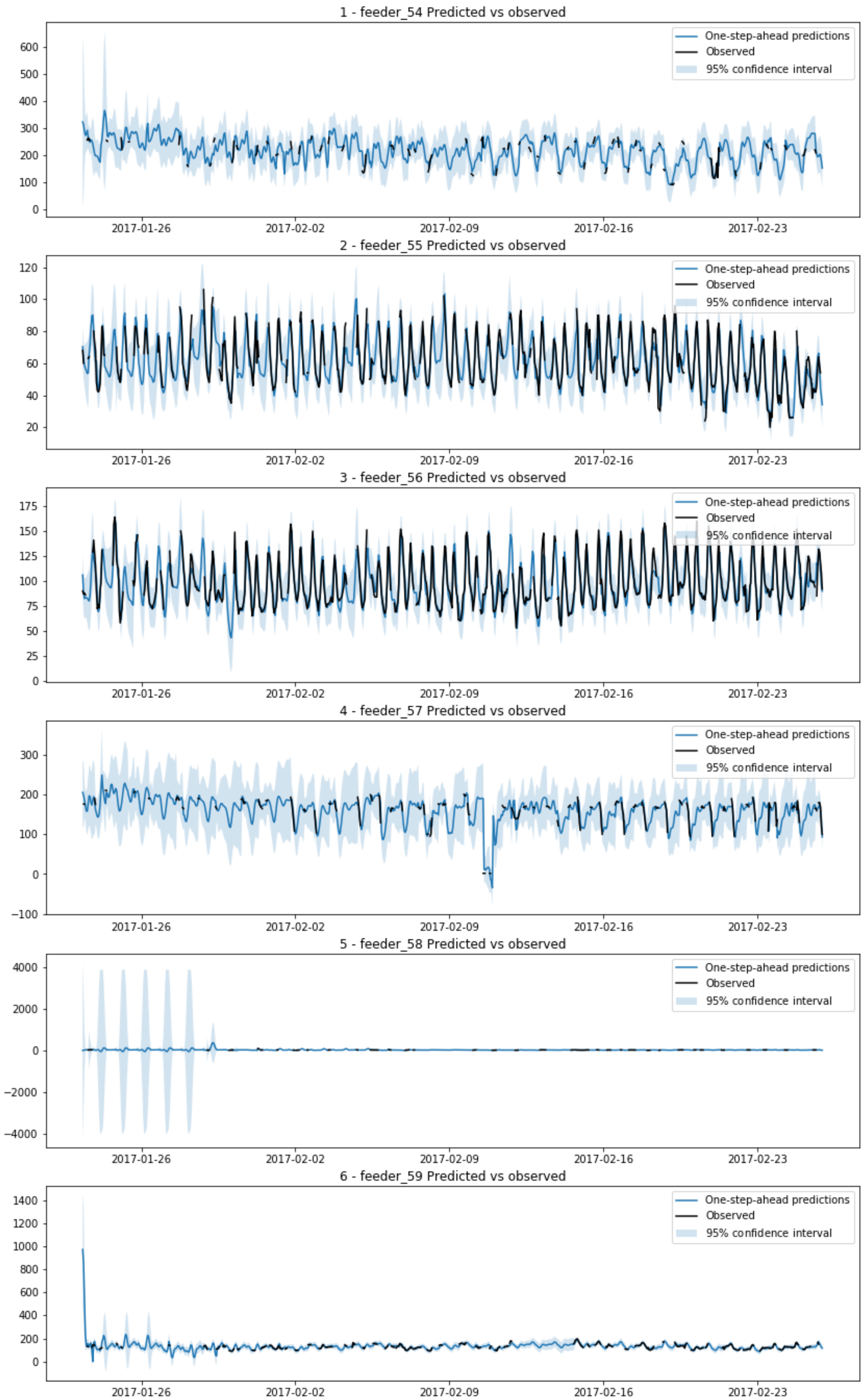


Figure 9.10: Observations compared to fitted values from structural models 54-59

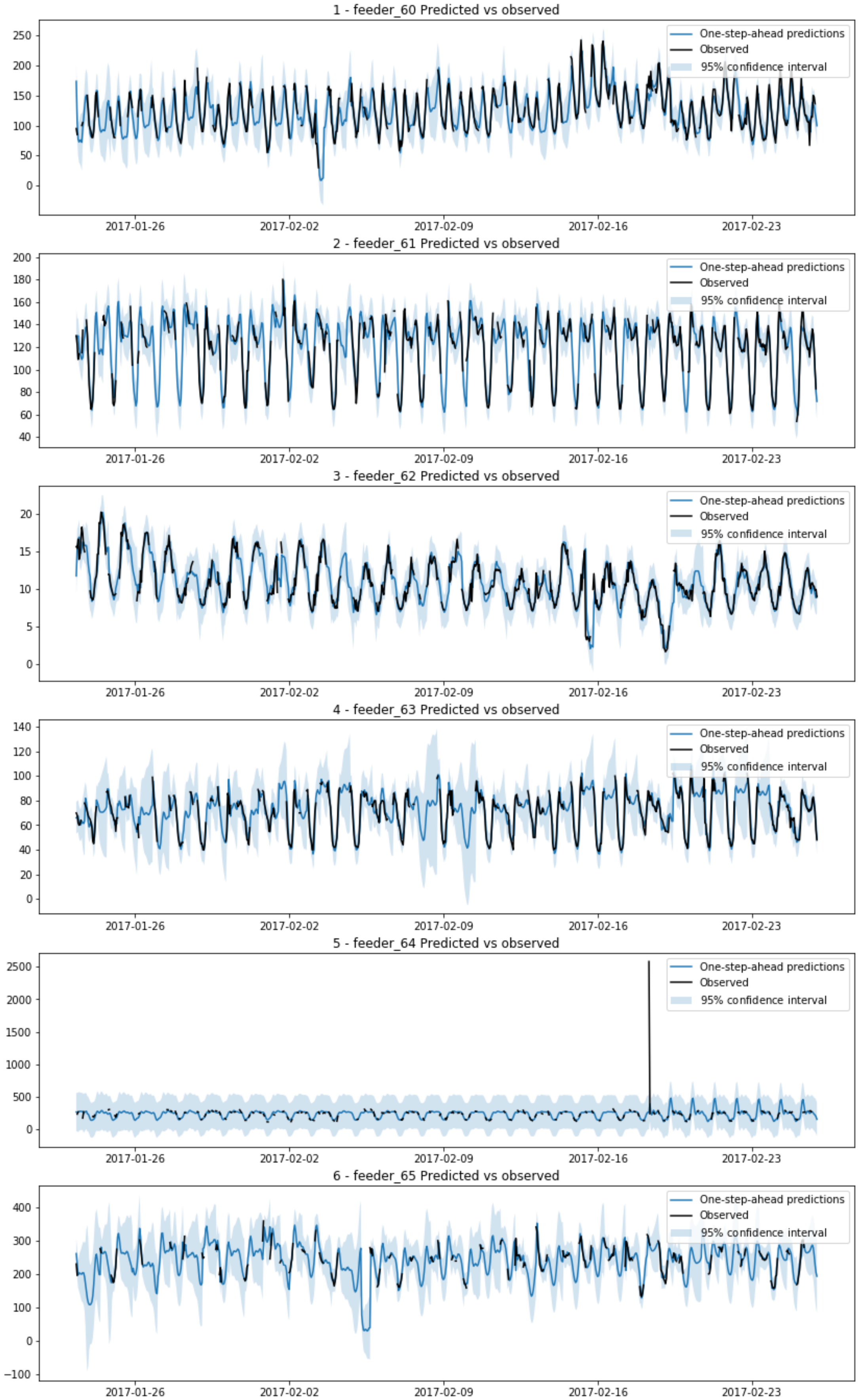


Figure 9.11: Observations compared to fitted values from structural models 60-65

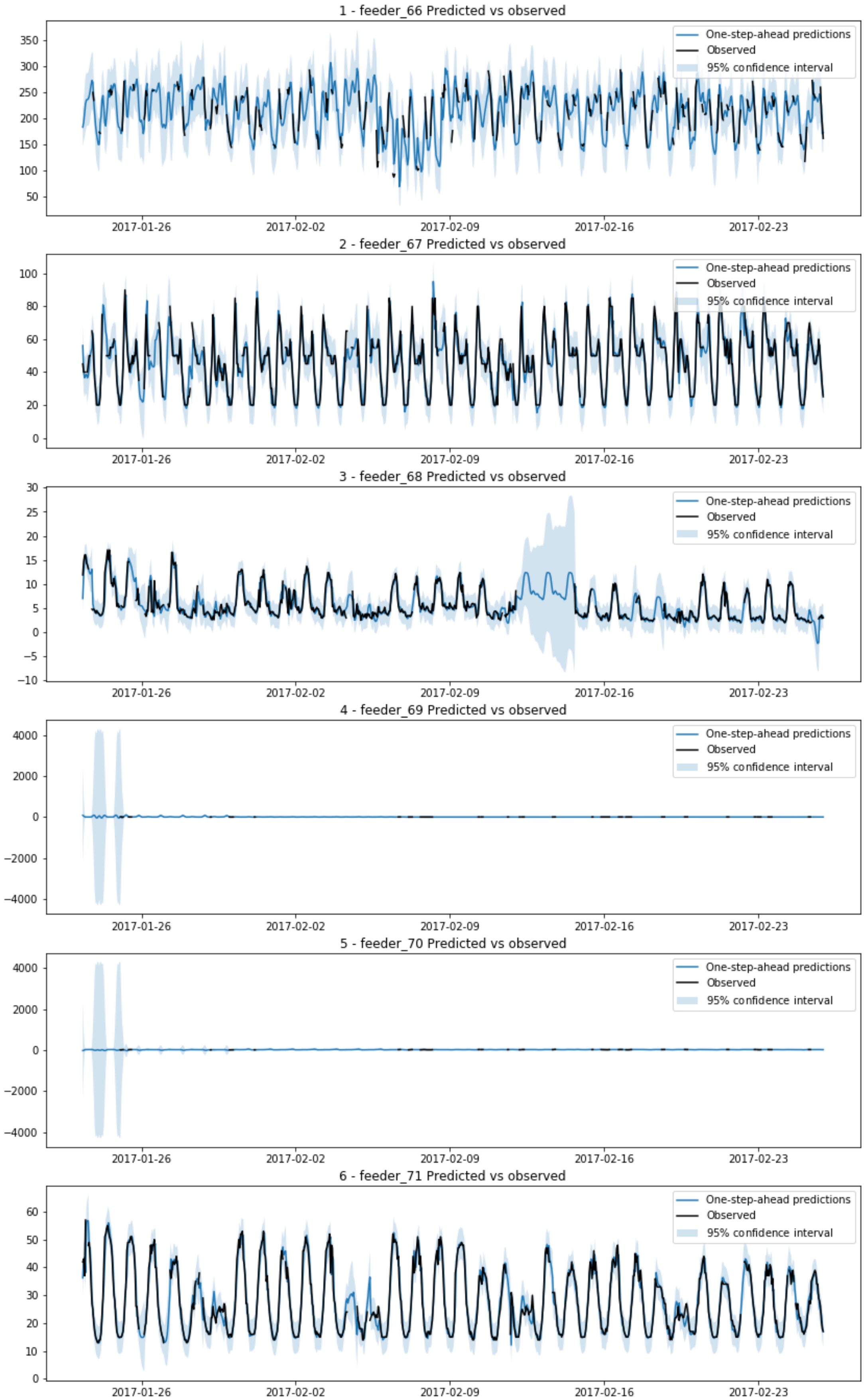


Figure 9.12: Observations compared to fitted values from structural models 65-71

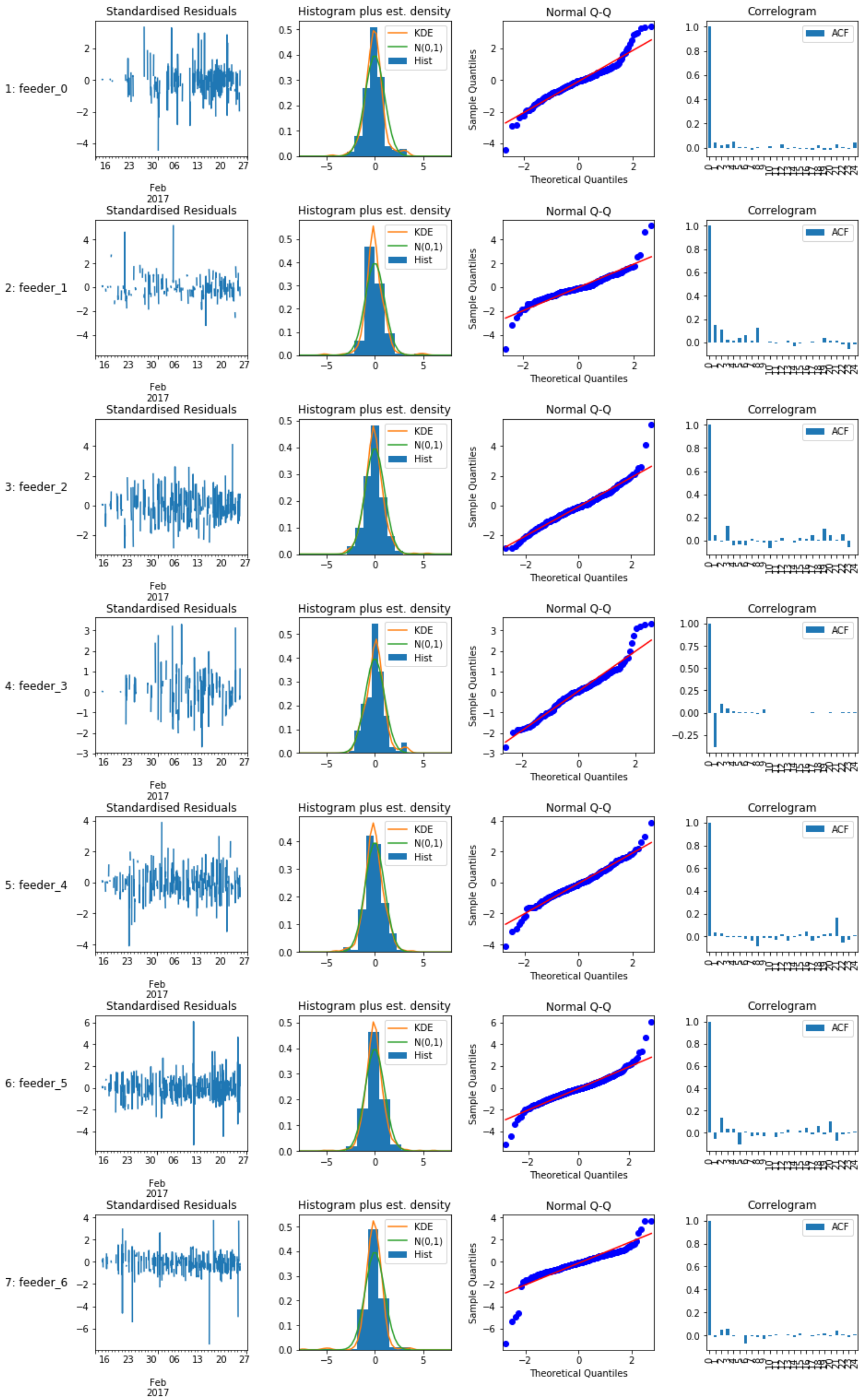


Figure 9.13: Structural model diagnostics 0-6

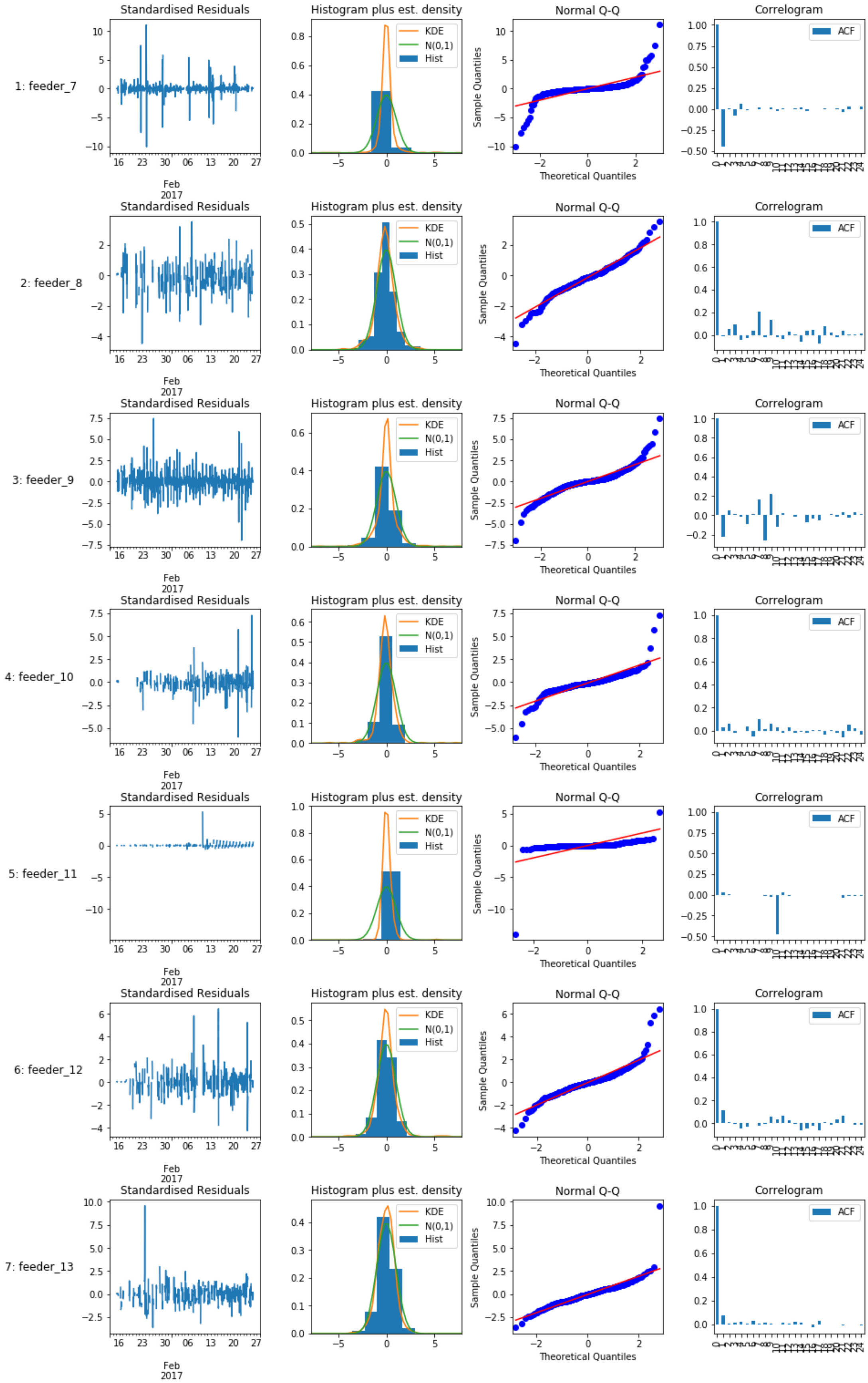


Figure 9.14: Structural model diagnostics 7-13

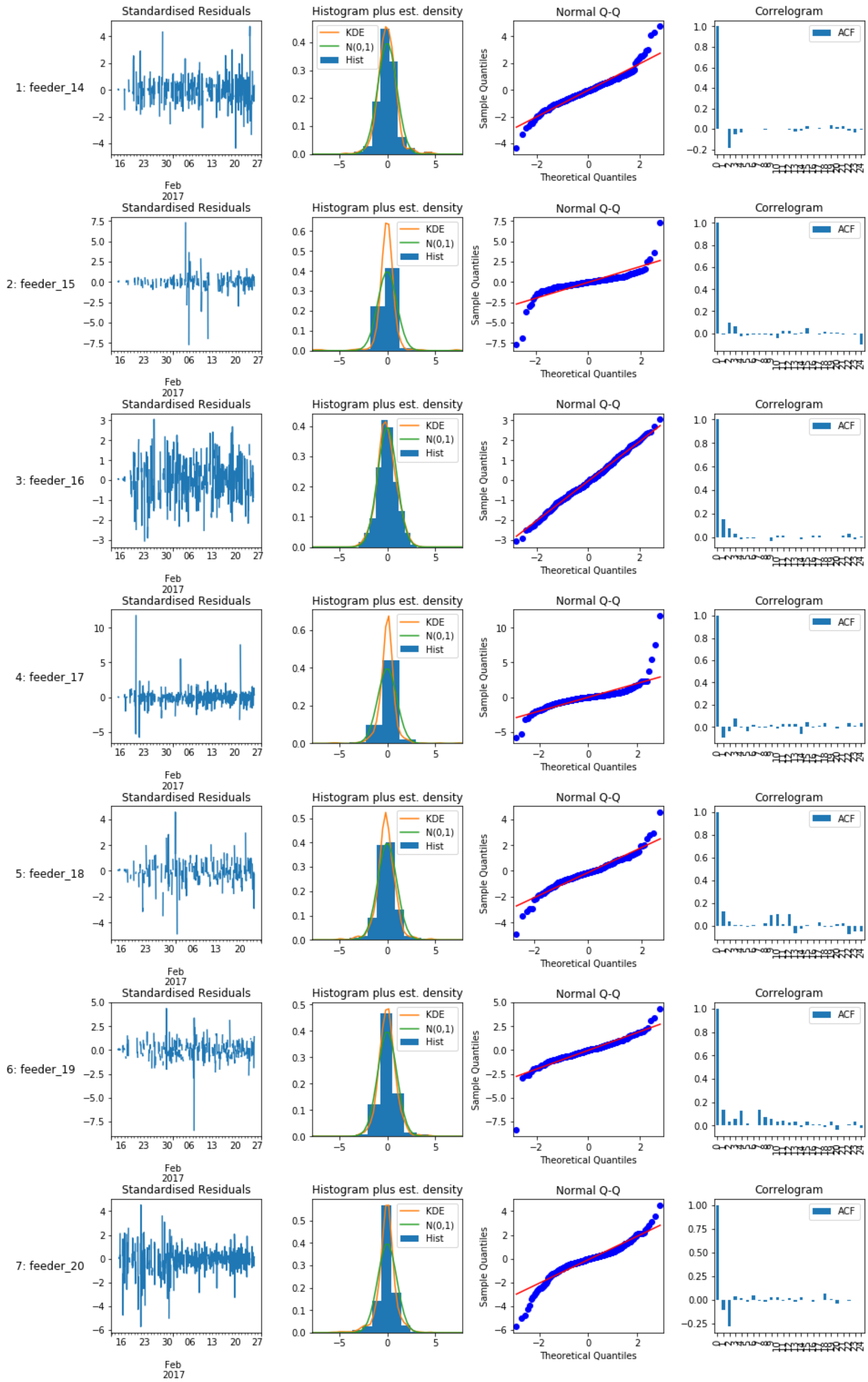


Figure 9.15: Structural model diagnostics 14-20

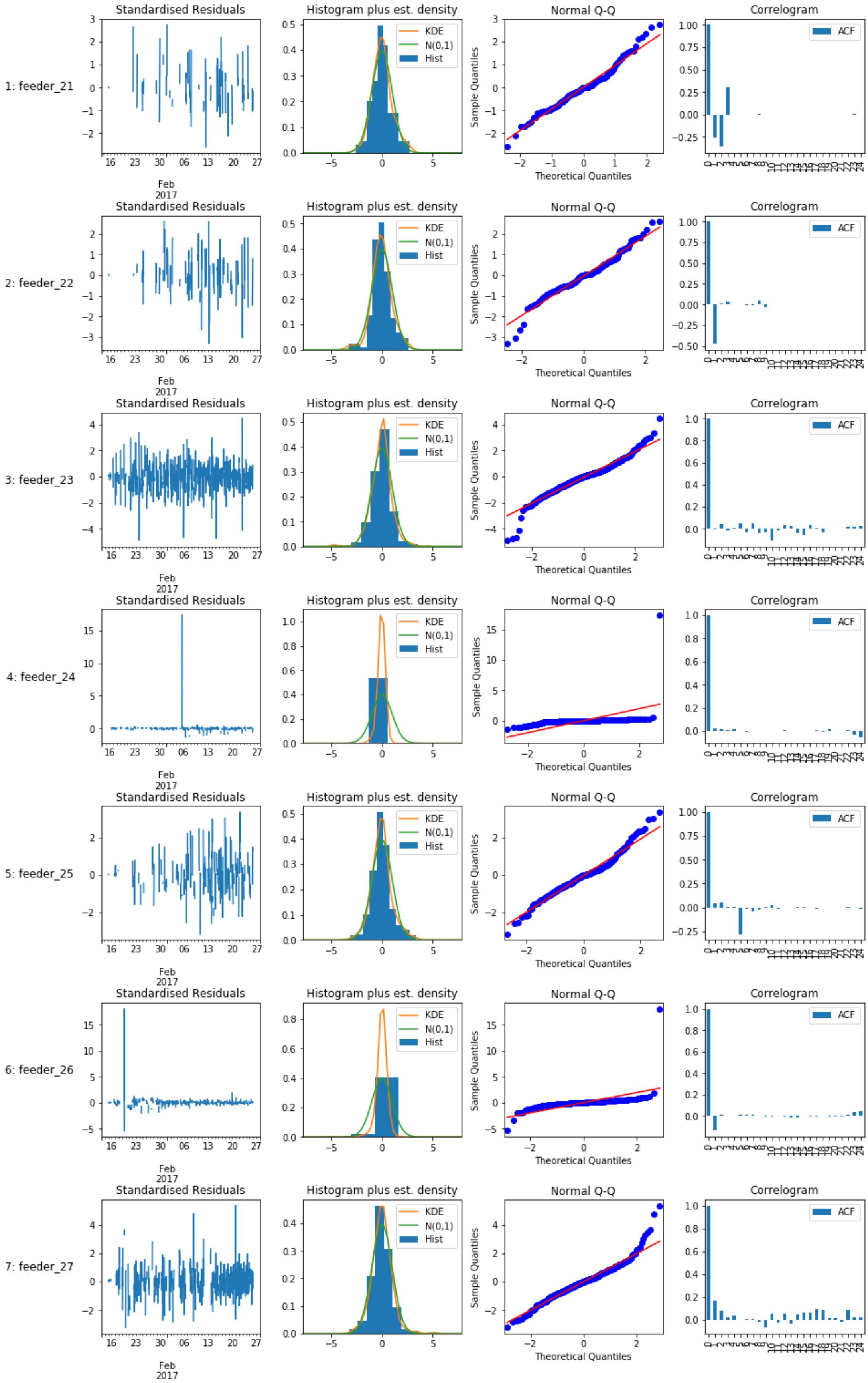


Figure 9.16: Structural model diagnostics 21-27

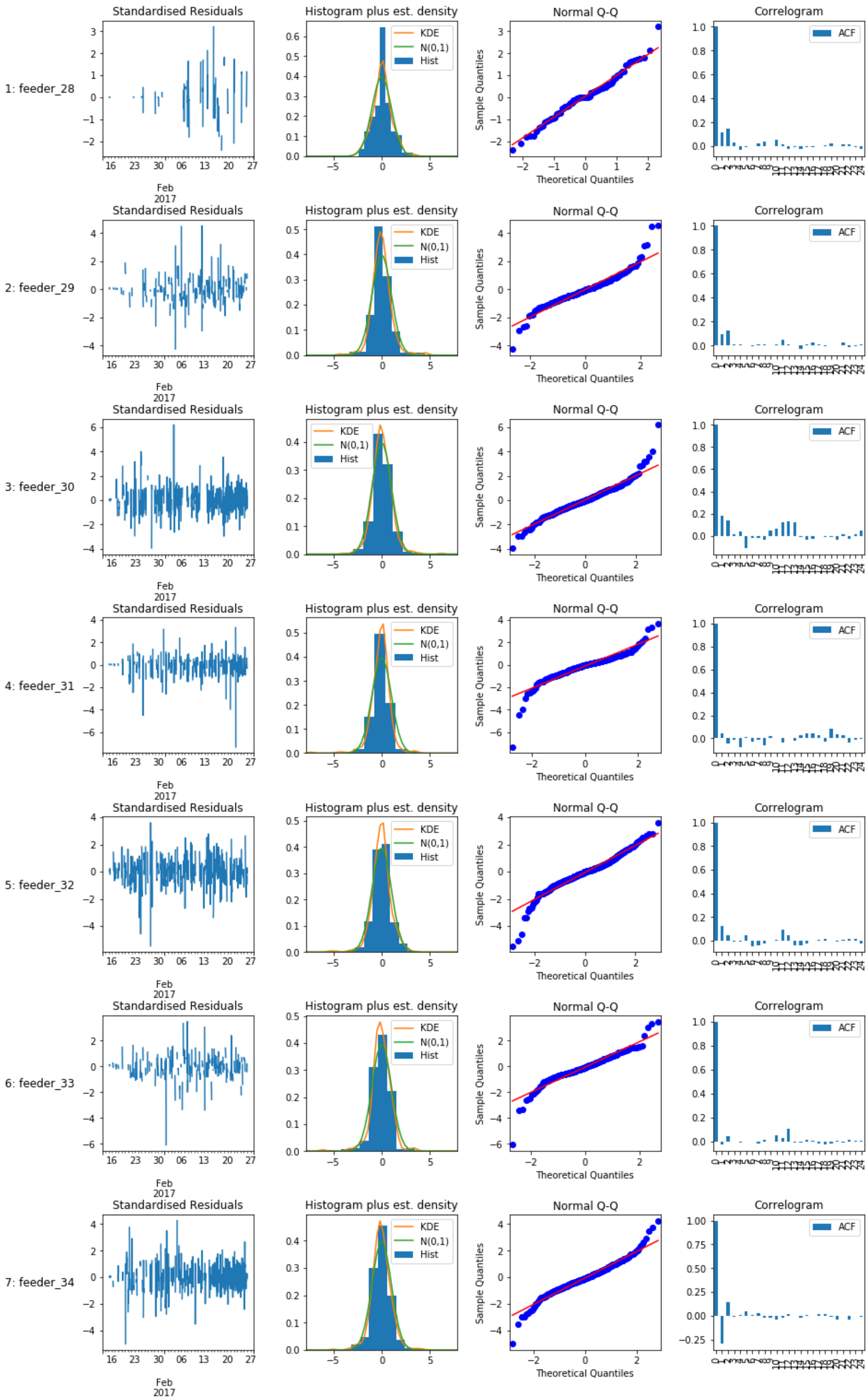


Figure 9.17: Structural model diagnostics 28-34

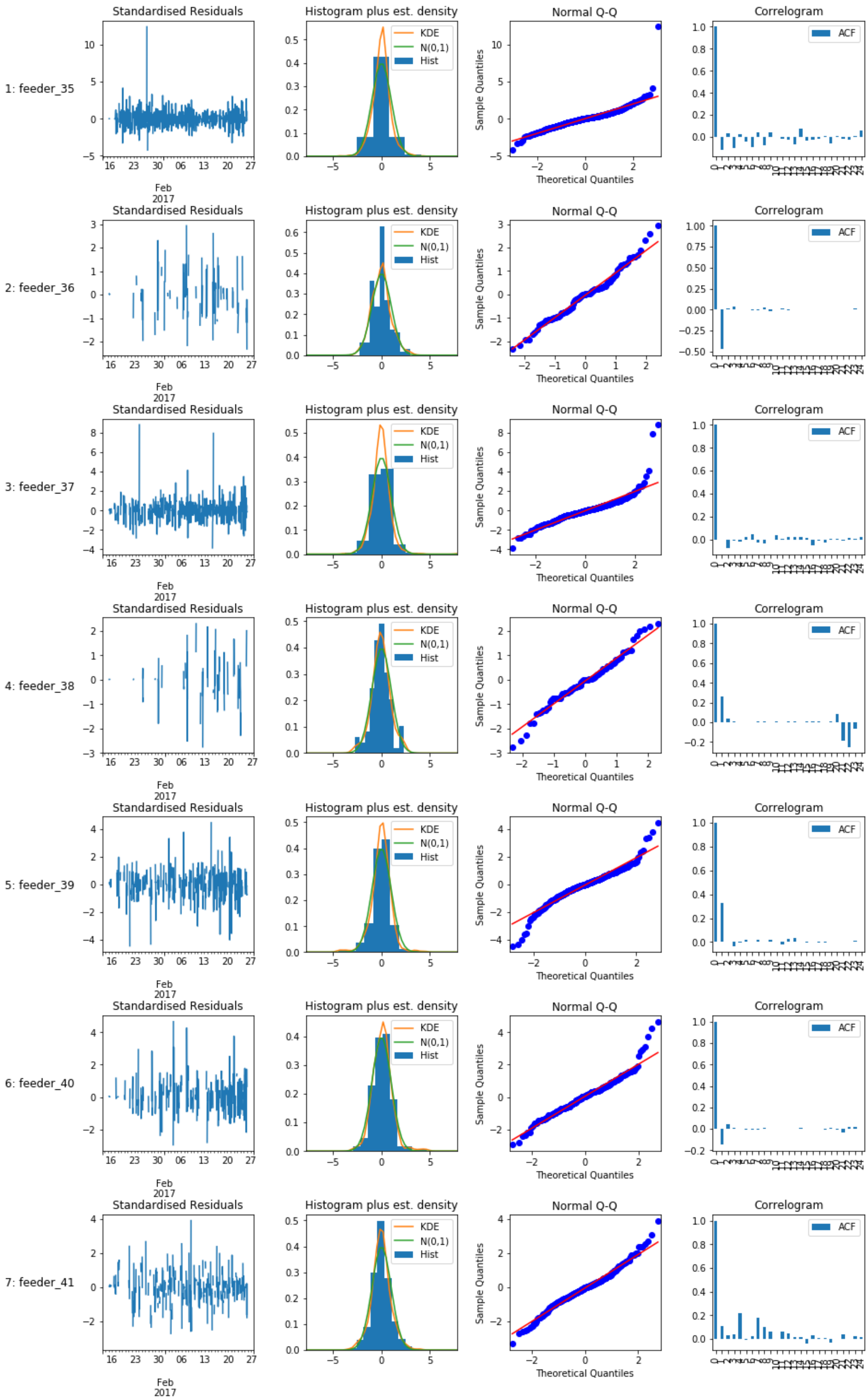


Figure 9.18: Structural model diagnostics 35-41

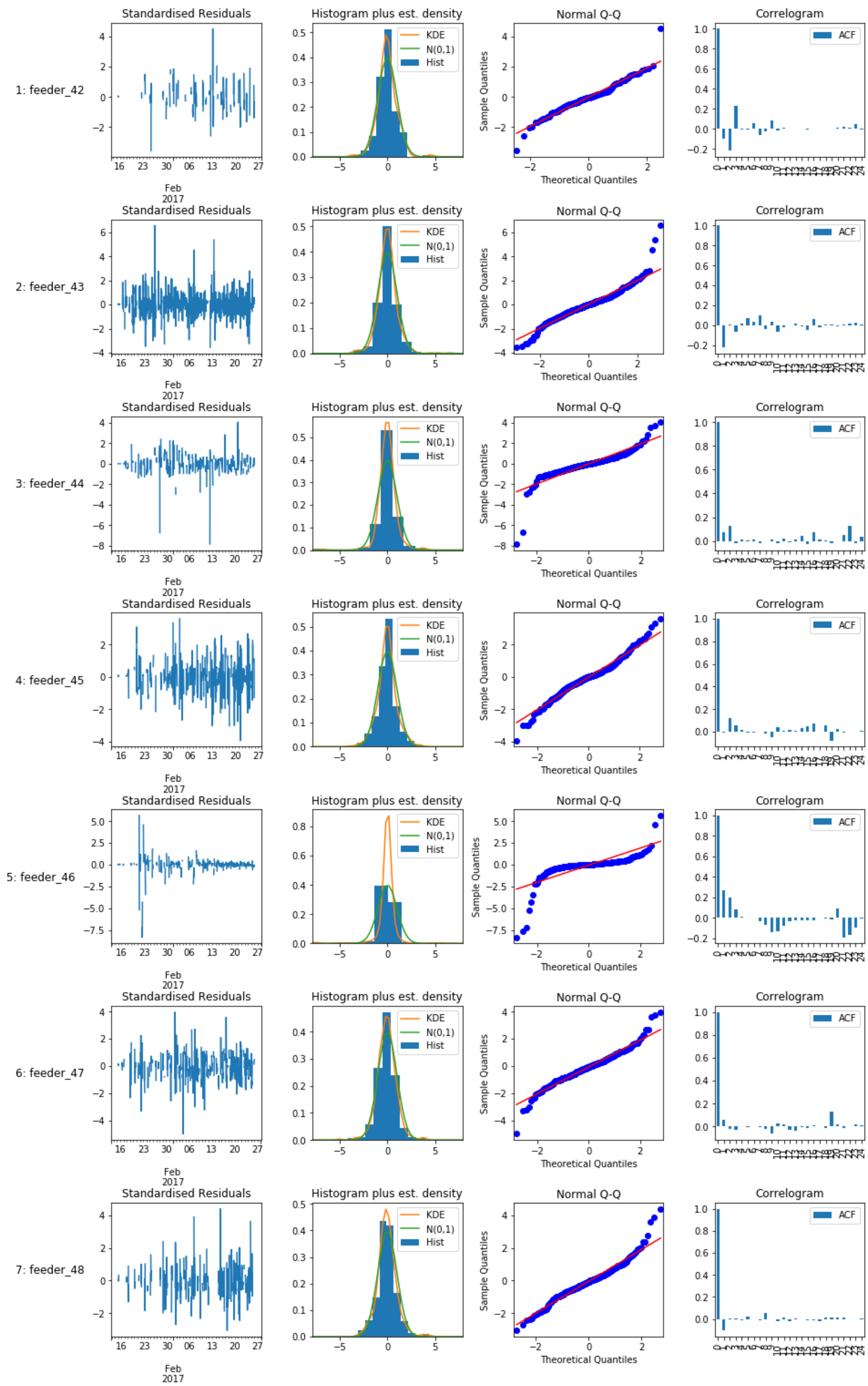


Figure 9.19: Structural model diagnostics 42-48

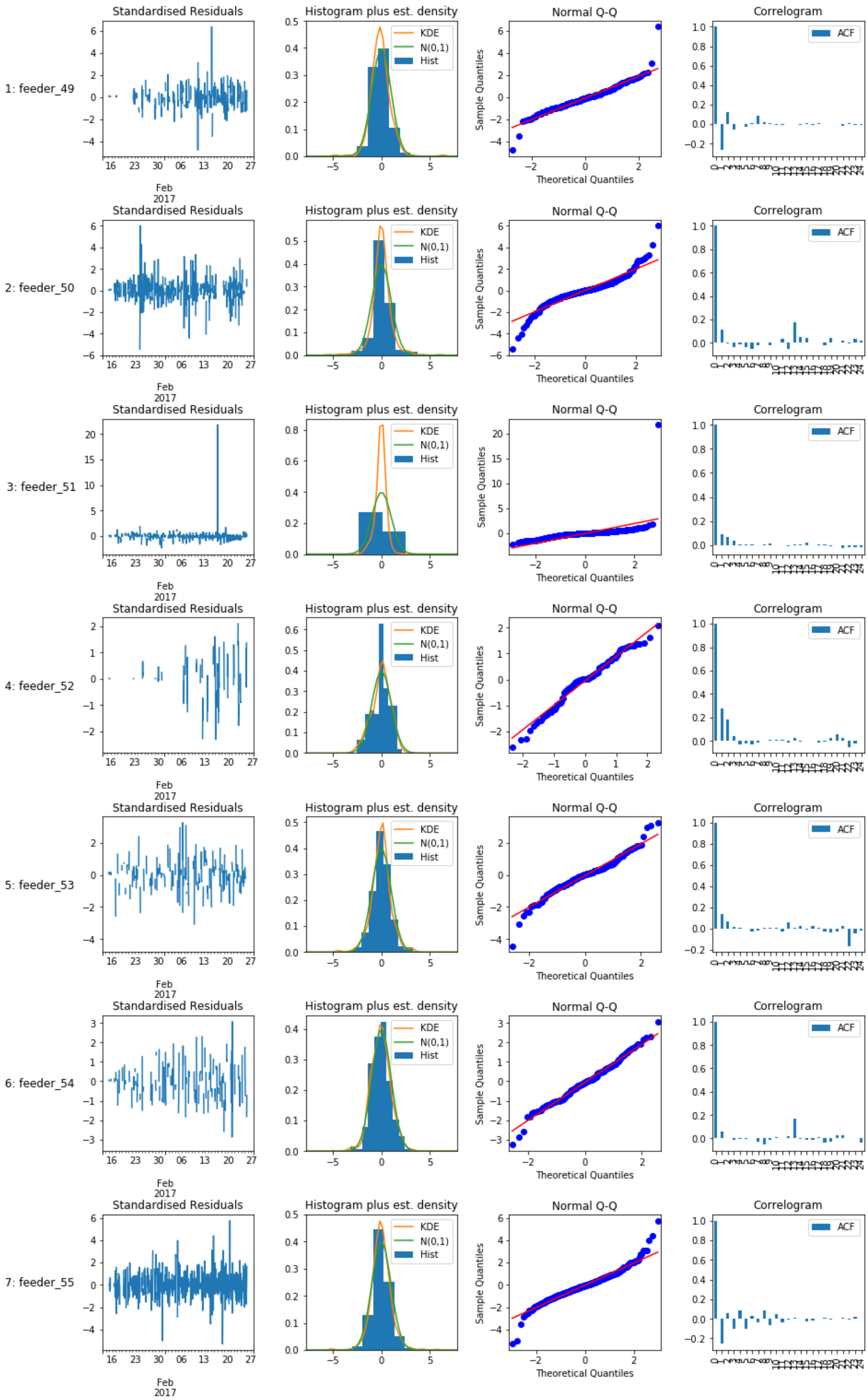


Figure 9.20: Structural model diagnostics 49-55

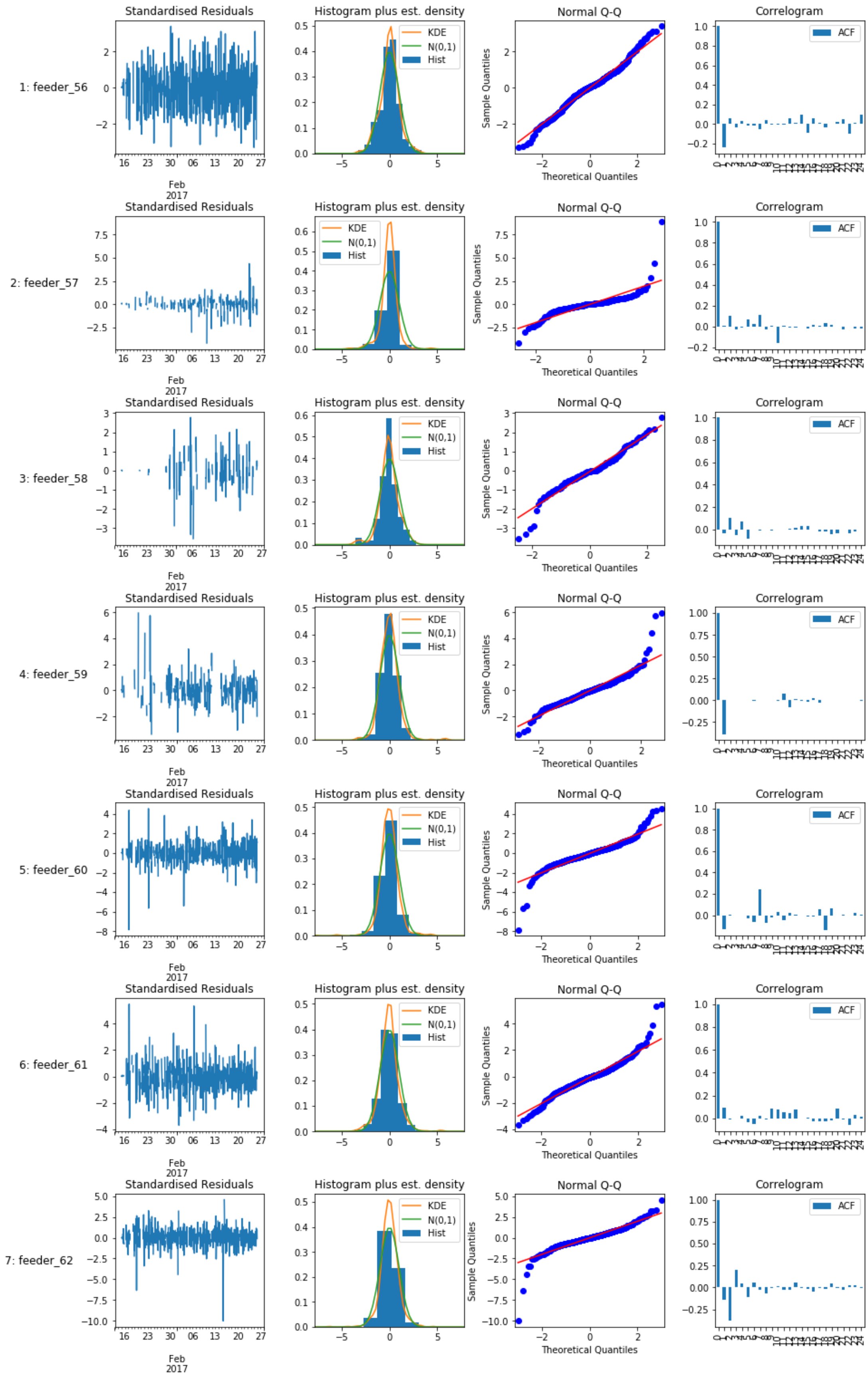


Figure 9.21: Structural model diagnostics 56-62

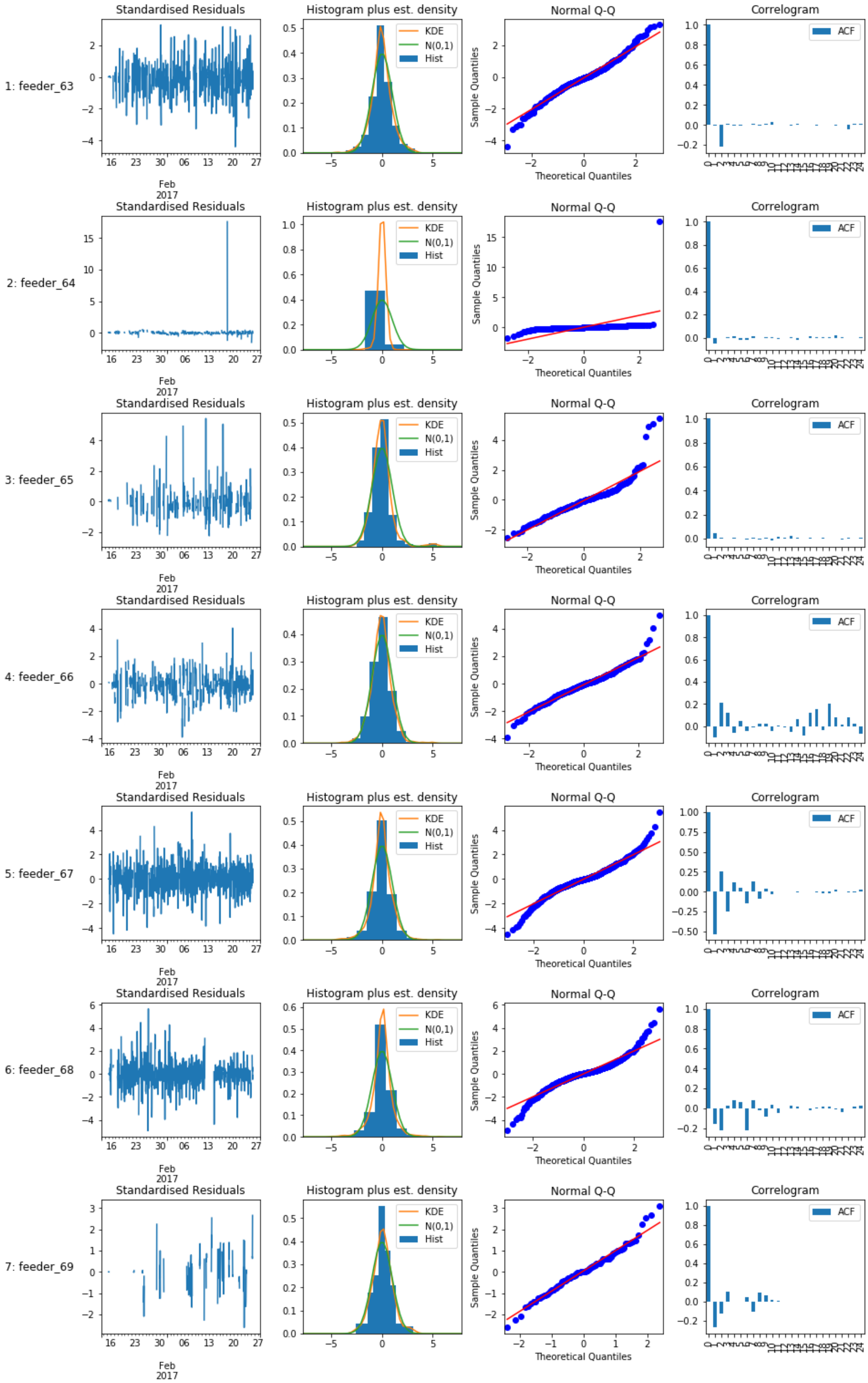


Figure 9.22: Structural model diagnostics 63-69

Chapter 10

Convexity of Myopic FSP

The stochastic knapsack problem can be continuously relaxed by taking $\mathbf{x_t} \in [0,1]^n$. The convexity of this continuous relaxation can be shown in two ways.

There is a theorem which says that stochastic programs with linear penalties for constraint violation and where the probability distributions have finite second moments are convex[8]. This is the approach taken in [40] to establish the convexity of the stochastic knapsack problem with random weights.

An alternative approach is to use elementary properties of convex functions to show that the objective function of the stochastic knapsack problem is concave providing that $\mu_t(\boldsymbol{x}_t)$ and $\sqrt{\sigma_t^2(\boldsymbol{x}_t)}$ are convex. $\mu_t(\boldsymbol{x}_t)$ is linear and $\sqrt{\sigma_t^2(\boldsymbol{x}_t)}$ is a norm and therefore convex. A proof of this is given in this appendix. The objective function of the multi-period feeder scheduling problem is the sum of the single feeder functions and is therefore the sum of concave functions and thus also concave.

Consider a normally distributed random variable D with a mean and standard deviation given by functions of $\boldsymbol{x}$, i.e. $D(\boldsymbol{x}) \sim N(\mu(\boldsymbol{x}), \sigma(\boldsymbol{x}))$

Next consider a function of D that takes the value D, if D is greater than a constant C and 0 otherwise. This is the random variable given by:

$(D(\boldsymbol{x}) - C)^+ = \max\{0, D(\boldsymbol{x} - C\}$

We wish to show that $\mathbb{E}[(D(\boldsymbol{x}) - C)^+]$ is a convex function of $\boldsymbol{x}$ if $\mu(\boldsymbol{x})$ and $\sigma(\boldsymbol{x})$ are convex.

It can be shown that:

$\mathbb{E}[(D(\boldsymbol{x}) - C)^+] = L\left(\frac{C-\mu_t(\boldsymbol{x}_t)}{\sigma_t(\boldsymbol{x}_t)}\right)\sigma_t(\boldsymbol{x}_t)$

where $L(z) = \phi(z) - z(1 - \Phi(z))$

and ϕ and $\Phi(z)$ are the PDF and CDF of the standard normal distribution with $\mu = 0$ and $\sigma = 1$.

We first show :

(1), $L(z)$ is convex and non-increasing

Proof:

1st derivative:

$$L'(z) = \phi'(z) - 1 + z\Phi'(z) + \Phi(z)$$

$$-z\phi(z) - 1 + z\phi(z) + \Phi(z) = \Phi(z) - 1$$

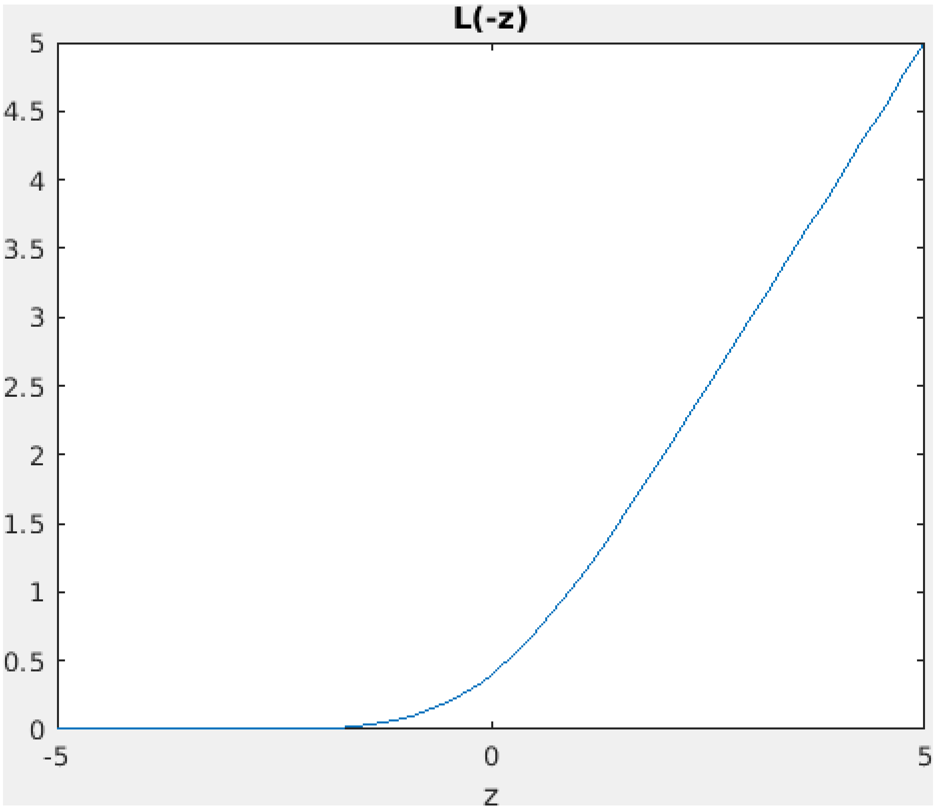
2nd derivative:

$$L''(z) = \Phi'(z) = \phi(z)$$

$\phi(z) \geq 0 \forall z$, therefore $L(z)$ is a convex function

It follows straightforwardly that:

(2), $L(-z)$ is convex and non-decreasing



Next we show:

(3), $\sigma L(\frac{C-\mu}{\sigma})$ is convex

Proof:

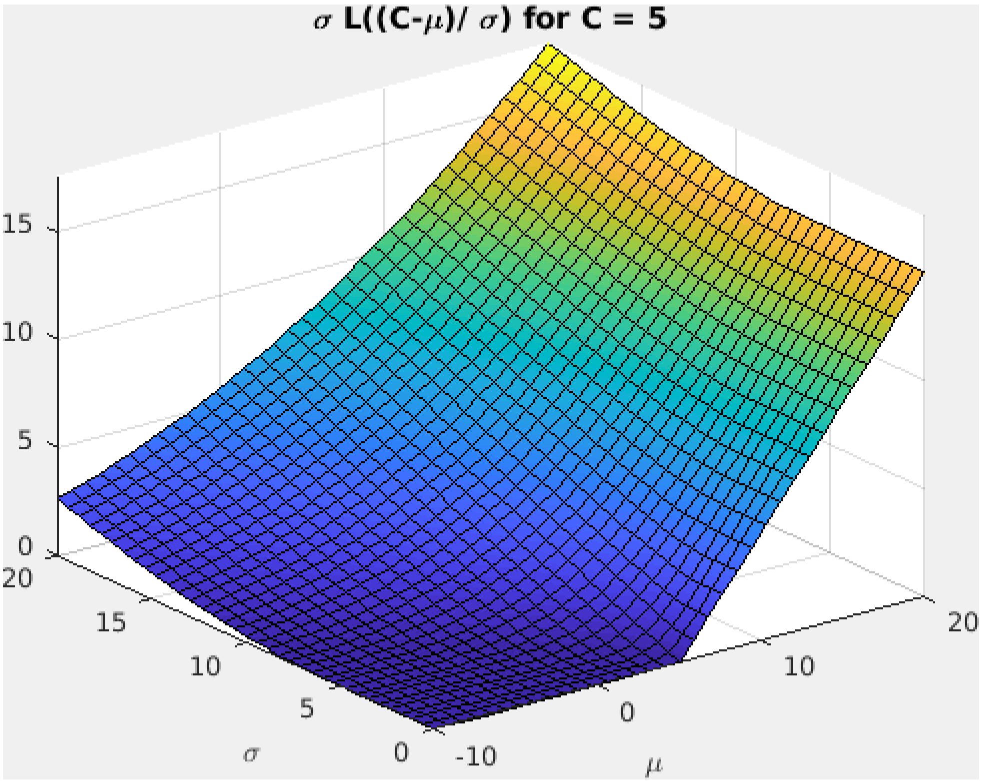
We use the following result (see [9] pp89):

The *perspective* of a function $f : \mathbf{R}^n \to \mathbf{R}$ is the function $g : \mathbf{R}^n \times \mathbf{R} \to \mathbf{R}$,

$g(x,t) = tf(x/t), \mathbf{dom} g = \{(x,t)|x/t \in \mathbf{dom} f, t > 0\}$

g is convex if f is convex.

If we take $-z = \mu - C$ it follows from (2) that $L(C-\mu)$ is convex and non-decreasing in μ. Therefore it's perspective $\sigma L(\frac{C-\mu}{\sigma})$ is convex.



(4a), $\sigma L(\frac{C-\mu}{\sigma})$ is non-decreasing in μ

Because $\sigma \geq 0$ this follows straightfowardly from (2)

(4b), $\sigma L(\frac{C-\mu}{\sigma})$ is non-decreasing in σ

We have previously shown that $L'(z) = \Phi(z) - 1$, therefore:

$$\frac{\partial}{\partial \sigma} \sigma L\Big(\frac{C-\mu}{\sigma}\Big) = L\Big(\frac{C-\mu}{\sigma}\Big) + \sigma\Big(\Phi\Big(\frac{C-\mu}{\sigma}\Big) - 1\Big)\Big(-\frac{C-\mu}{\sigma^2}\Big)$$

$$= L\Big(\frac{C-\mu}{\sigma}\Big) + \Big(1 - \Phi\Big(\frac{C-\mu}{\sigma}\Big)\Big)\Big(\frac{C-\mu}{\sigma}\Big)$$

$$= \phi\left(\frac{C-\mu}{\sigma}\right) - \left(\frac{C-\mu}{\sigma}\right)\left(1-\Phi\left(\frac{C-\mu}{\sigma}\right)\right) + \left(1-\Phi\left(\frac{C-\mu}{\sigma}\right)\right)\left(\frac{C-\mu}{\sigma}\right)$$

$$\frac{\partial}{\partial\sigma}\sigma L\left(\frac{C-\mu}{\sigma}\right) = \phi\left(\frac{C-\mu}{\sigma}\right)$$

The required result follows since $\phi(z)$ is positive.

Conclusion

We use the following result (see [9] pp86):

Suppose

$$f(x) = h(g(x)) = h(g_1(x),...,g_2(x))$$

with $h : \mathbf{R}^k \to \mathbf{R}$, $g : \mathbf{R}^n \to \mathbf{R}$

f is convex if h is convex, h is nondecreasing in each argument and g_i are convex

---

Define $h(\mu(\boldsymbol{x}),\sigma(\boldsymbol{x})) = L(\frac{C-\mu(\boldsymbol{x})}{\sigma(\boldsymbol{x})})$

By assumption $\mu(\boldsymbol{x})$ and $\sigma(\boldsymbol{x})$ are convex.

This combined with (3), (4a) and (4b) proves the required result.

Chapter 11

Autocorrelation of Demand Forecasts in the Feeder Scheduling Problem

Let $\mathcal{T}$ denote a subset of $\mathcal{P}$ containing all the periods in a given decision interval. The expected rewards minus penalties given decisions $\boldsymbol{x}_t \forall t \in \mathcal{T}$ is given by.

$\sum_{t\in\mathcal{T}} (\boldsymbol{r} \circ \boldsymbol{\mu}_t)^T \boldsymbol{x}_t - p\mathbb{E}[\sum_{t\in\mathcal{T}} (D_t(\boldsymbol{x}_t) - C)^+]$

In order to evaluate the second term, the expected penalties, we need to use the joint density of $D_t(\boldsymbol{x}_t) \forall t \in \mathcal{T}$ and integrate with respect to demand at all time periods.

To see how to do this consider a simple case wtih two periods. Suppose the demand in one period is X and another period is Y and that we are penalised by p per unit of excess demand in each period. The expected penalties is:

$p\mathbb{E}[(X-C)^+ + (Y-C)^+]$

Let $f(X,Y)$ be the joint density function of X and Y. The expectation is given by the double integral:

$$p\int (X-C)^+ + (Y-C)^+ f(X,Y)\mathrm{d}(X,Y)$$

Using the linearity of integration this is equal to:

$$p\int (X-C)^+ f(X,Y)d(X,Y) + p\int (Y-C)^+) f(X,Y)\mathrm{d}(X,Y)$$

Fubini's theorem says that under very mild conditions we can replace the double integral with an iterated integral and therefore also change the order of integration. To apply this theorem it is required that the double integral is finite if its integrand is replaced by its absolute value. In our case the integrand of each integral is equal to its absolute value because the integrand is always positive.

$p\int (X-C)^+ f(X,Y)\mathrm{d}(X,Y)$ is finite.

$p\int (X-C)^+ f(X,Y)\mathrm{d}(X,Y) = p\int\int (X-C)^+ f(X,Y)\mathrm{d}Y\mathrm{d}X = p\int (X-C)^+ \int f(X,Y)\mathrm{d}Y\mathrm{d}X$

$\int f(X,Y)\mathrm{d}Y = f(X)$ is the marginal density of X, which is a normal PDF, hence

$p\int (X-C)^{+}\int f(X,Y)\mathrm{d}Y\mathrm{d}X = p\int (X-C)^{+} f(X)\mathrm{d}X = p\mathbb{E}[(X-C)^{+}]$

Therefore $p\mathbb{E}[(X-C)^{+} + (Y-C)^{+}] = p\mathbb{E}[(X-C)^{+}] + p\mathbb{E}[(Y-C)^{+}]$

It follows that even in the case where demand is correlated between time points, we can treat the problem as the sum of several stochastic knapsack problems.

Bibliography

[1] Ahlborg, H., Boräng, F., Jagers, S. C., and Söderholm, P. Provision of electricity to African households: The importance of democracy and institutional quality. *Energy Policy*, 87:125–135, dec 2015. ISSN 0301-4215. doi: 10.1016/J.ENPOL.2015.09.002. URL https://www.sciencedirect.com/science/article/pii/S0301421515300884.

[2] Akinlo, A. Electricity consumption and economic growth in Nigeria: Evidence from cointegration and co-feature analysis. *Journal of Policy Modeling*, 31(5):681–693, sep 2009. doi: 10.1016/J.JPOLMOD.2009.03.004. URL https://www.sciencedirect.com/science/article/pii/S0161893809000180.

[3] Aliyu, A. K., Modu, B., and Tan, C. W. A review of renewable energy development in Africa: A focus in South Africa, Egypt and Nigeria. *Renewable and Sustainable Energy Reviews*, 81:2502–2518, jan 2018. doi: 10.1016/J.RSER.2017.06.055. URL https://www.sciencedirect.com/science/article/pii/S1364032117309887.

[4] Aliyu, A. S., Ramli, A. T., and Saleh, M. A. Nigeria electricity crisis: Power generation capacity expansion and environmental ramifications. *Energy*, 61:354–367, nov 2013. ISSN 03605442. doi: 10.1016/j.energy.2013.09.011. URL http://www.sciencedirect.com/science/article/pii/S0360544213007627.

[5] Anderson, B. D. O. *Optimal filtering*. Mineola, N.Y. : Dover, Mineola, N.Y., 2005.

[6] Arango, S., Dyner, I., and Larsen, E. R. Lessons from deregulation: Understanding electricity markets in South America. *Utilities Policy*, 14(3):196–207, 2006. ISSN 09571787. doi: 10.1016/j.jup.2006.02.001. URL http://www.sciencedirect.com/science/article/pii/S0957178706000130.

[7] Bernard, J.-T. and Roland, M. Load management programs, cross-subsidies and transaction costs: the case of self-rationing. *Resource and Energy Economics*, 22(2):161–188, 2000. doi: 10.1016/S0928-7655(99)00018-4.

[8] Birge, J. R. and Louveaux, F. *Introduction to Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer New York, New York, NY, 2011. ISBN 978-1-4614-0236-7. doi: 10.1007/978-1-4614-0237-4. URL http://link.springer.com/10.1007/978-1-4614-0237-4.

[9] Boyd, S. P. *Convex optimization*. Cambridge University Press, Cambridge, 2004. ISBN 9780521833783.

[10] Bukhsh, W. A., Grothey, A., McKinnon, K. I. M., and Trodden, P. A. Local Solutions of the Optimal Power Flow Problem. *IEEE Transactions on Power Systems*, 28(4):4780–4788, nov 2013. ISSN 0885-8950. doi: 10.1109/TPWRS.2013.2274577. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6581918.

[11] Caivano, M., Harvey, A., and Luati, A. Robust time series models with trend and seasonal components. *SERIEs*, 7(1):99–120, mar 2016. ISSN 1869-4187. doi: 10.1007/s13209-015-0134-1. URL http://link.springer.com/10.1007/s13209-015-0134-1.

[12] Cooley, T. F. and Prescott, E. C. An Adaptive Regression Model. *International Economic Review*, 14(2):364, jun 1973. doi: 10.2307/2525926. URL https://www.jstor.org/stable/2525926?origin=crossref.

[13] Darbellay, G. A. and Slama, M. Forecasting the short-term demand for electricity: Do neural networks stand a better chance? *International Journal of Forecasting*, 16 (1):71–83, jan 2000. doi: 10.1016/S0169-2070(99)00045-X. URL https://www.sciencedirect.com/science/article/pii/S016920709900045X.

[14] Ding Xu, D. and Girgis, A. A. Optimal load shedding strategy in power systems with distributed generation. In *2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.01CH37194)*, volume 2, pages 788–793. IEEE, 2001. ISBN 0-7803-6672-7. doi: 10.1109/PESW.2001.916959. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=916959.

[15] Dordonnat, V. and Ooms, M. Dynamic factors in periodic time-varying regressions with an application to hourly electricity load modelling. *Computational Statistics & Data Analysis*, 56(11):3134–3152, nov 2012. doi: 10.1016/J.CSDA.2011.04.002. URL https://www.sciencedirect.com/science/article/pii/S0167947311001290.

[16] Doucet, J. A., Jo Min, K., Roland, M., and Strauss, T. Electricity rationing through a two-stage mechanism. *Energy Economics*, 18(3):247–263, 1996. doi: 10.1016/0140-9883(96)00014-X.

[17] Durbin, J. and Koopman, S. J. *Time series analysis by state space methods*. Oxford : Oxford University Press, Oxford, 2nd ed.. edition, 2012.

[18] Eberhard, A. A., Foster, V., and Briceno-Garme. Africa - Underpowered : the state of the power sector in Sub-Saharan Africa, 2008. URL http://documents.worldbank.org/curated/en/142991468006934762/Africa-Underpowered-the-state-of-the-power-sector-in-Sub-Saharan-Africa.

[19] ESMAP. BEYOND CONNECTIONS Energy Access Redefined. Technical report, 2015.

[20] Farquharson, D., Jaramillo, P., and Samaras, C. Sustainability implications of electricity outages in sub-Saharan Africa. *Nature Sustainability*, 1(10):589–597, oct 2018. doi: 10.1038/s41893-018-0151-8. URL http://www.nature.com/articles/s41893-018-0151-8.

[21] Fernandes, T. S. P., Lenzi, J. R., and Mikilita, M. A. Load Shedding Strategies Using Optimal Load Flow With Relaxation of Restrictions. *IEEE Transactions on Power Systems*, 23(2):712–718, may 2008. doi: 10.1109/TPWRS.2008.919237. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4494595.

[22] Gratwick, K. N. and Eberhard, A. Demise of the standard model for power sector reform and the emergence of hybrid power markets. *Energy Policy*, 36(10):3948–3960, 2008. ISSN 03014215. doi: 10.1016/j.enpol.2008.07.021.

[23] Han, J., Lee, K., Lee, C., Choi, K. S., and Park, S. Robust optimization approach for a chance-constrained binary knapsack problem. *Mathematical Programming*, 157(1):277–296, may 2016. ISSN 14364646. doi: 10.1007/s10107-015-0931-0. URL http://link.springer.com/10.1007/s10107-015-0931-0.

[24] Harvey, A. C. and Phillips, G. D. A. The Estimation of Regression Models with Time-Varying Parameters. In *Games, Economic Dynamics, and Time Series Analysis*, pages 306–321. Physica-Verlag HD, Heidelberg, 1982. doi: 10.1007/978-3-662-41533-7_18. URL http://link.springer.com/10.1007/978-3-662-41533-7{_}18.

[25] Harvey, A. and Luati, A. Filtering With Heavy Tails. *Journal of the American Statistical Association*, 109(507):1112–1122, jul 2014. ISSN 1537274X. doi: 10.1080/01621459.2014.887011. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.2014.887011.

[26] Harvey, A. C. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge : Cambridge University Press, 1990.

[27] Heggie, A., Eager, D., McKinnon, K., and Van Der Weijde, A. H. Power rationing in a long-term power shortage. *Energy Policy*, 121, 2018. ISSN 03014215. doi: 10.1016/j.enpol.2018.06.025.

[28] Hiriart-Urruty, J.-B. and Lemaréchal, C. *Convex Analysis and Minimization Algorithms II : Advanced Theory and Bundle Methods*. Springer Berlin Heidelberg, 1993. ISBN 9783642081620.

[29] Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., and Hyndman, R. J. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *International Journal of Forecasting*, 32(3):896–913, jul 2016. doi: 10.1016/J.IJFORECAST.2016.02.001. URL https://www.sciencedirect.com/science/article/pii/S0169207016000133{#}f000020.

[30] Hunt, S. *Making competition work in electricity*. J. Wiley, New York, 2002.

[31] Hyndman, R. J. *Forecasting with exponential smoothing the state space approach*. Berlin : Springer, Berlin, 2008.

[32] Infield, D. G. and Hill, D. C. Optimal smoothing for trend removal in short term electricity demand forecasting. *IEEE Transactions on Power Systems*, 13(3):1115–1120, 1998. ISSN 08858950. doi: 10.1109/59.709108. URL http://ieeexplore.ieee.org/document/709108/.

[33] International Finance Corporation. From Gap to Opportunity: Business Models for Scaling Up Energy Access. Technical report, International Finance Corporation, 2012.

[34] Iyke, B. N. Electricity consumption and economic growth in Nigeria: A revisit of the energy-growth debate. *Energy Economics*, 51:166–176, sep 2015. ISSN 01409883. doi: 10.1016/j.eneco.2015.05.024. URL http://www.sciencedirect.com/science/article/pii/S0140988315001887.

[35] Kleywegt, A. J., Shapiro, A., and Homem-de Mello, T. The Sample Average Approximation Method for Stochastic Discrete Optimization. *SIAM Journal on Optimization*, 12(2):479–502, jan 2002. ISSN 1052-6234. doi: 10.1137/S1052623499363220. URL http://epubs.siam.org/doi/10.1137/S1052623499363220.

[36] Kosuch, S. and Lisser, A. Upper bounds for the 0-1 stochastic knapsack problem and a B&B algorithm. *Annals of Operations Research*, 176(1):77–93, apr 2010. ISSN 02545330. doi: 10.1007/s10479-009-0577-5. URL http://link.springer.com/10.1007/s10479-009-0577-5.

[37] Kumar, V., Gupta, I., and Gupta, H. O. An Overview of Cold Load Pickup Issues in Distribution Systems. *Electric Power Components and Systems*, 34(6):639–651, jun 2006. doi: 10.1080/15325000500419219. URL https://www.tandfonline.com/doi/full/10.1080/15325000500419219.

[38] Leou, R., Gaing, Z., Lu, C., Chang, B., and Cheng, C. Distribution system feeder cold load pickup model. *Electric Power Systems Research*, 36(3):163–168, mar 1996. doi: 10.1016/0378-7796(95)01027-0. URL https://www.sciencedirect.com/science/article/pii/0378779695010270.

[39] Marchand, M. G. Pricing power supplied on an interruptible basis. *European Economic Review*, 5(3):263–274, jan 1974. doi: 10.1016/0014-2921(74)90029-4. URL http://linkinghub.elsevier.com/retrieve/pii/0014292174900294.

[40] Merzifonluoğlu, Y., Geunes, J., and Romeijn, H. E. The static stochastic knapsack problem with normally distributed item sizes. *Mathematical Programming*, 134(2):459–489, sep 2012. ISSN 00255610. doi: 10.1007/s10107-011-0443-5. URL http://link.springer.com/10.1007/s10107-011-0443-5.

[41] Odumugbo, C. A. Natural gas utilisation in Nigeria: Challenges and opportunities. *Journal of Natural Gas Science and Engineering*, 2(6):310–316, dec 2010. ISSN 18755100. doi: 10.1016/j.jngse.2010.08.004. URL http://www.sciencedirect.com/science/article/pii/S1875510010000673.

[42] Oren, S. S. and Doucet, J. A. Interruption insurance for generation and distribution of electric power. *Journal of Regulatory Economics*, 2(1):5–19, mar 1990. doi: 10.1007/BF00139359. URL http://link.springer.com/10.1007/BF00139359.

[43] Oseni, M. O. An analysis of the power sector performance in Nigeria. *Renewable and Sustainable Energy Reviews*, 15(9):4765–4774, dec 2011. ISSN 13640321. doi: 10.1016/j.rser.2011.07.075. URL http://www.sciencedirect.com/science/article/pii/S1364032111003200.

[44] Rudnick, H. Market Restructuring In South America. *IEEE Power Engineering Review*, 18(6):3–6, jun 1998. ISSN 0272-1724. doi: 10.1109/MPER.1998.676632. URL http://ieeexplore.ieee.org/document/676632/.

[45] Sambo, A. S. and Garba, B. Electricity generation and the present challenges in the Nigerian power sector. *Journal of Energy and ...*, 2012. URL http://www.indiaenergycongress.in/montreal/library/pdf/70.pdf.

[46] Seabold, Skipper; Perktold, J. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, 2010.

[47] Showers, K. B. Electrifying africa: an environmental history with policy implications. *Geografiska Annaler: Series B, Human Geography*, 93(3):193–221, sep 2011. doi: 10.1111/j.1468-0467.2011.00373.x. URL https://www.tandfonline.com/doi/full/10.1111/j.1468-0467.2011.00373.x.

[48] Simon, D. *Optimal state estimation : Kalman, H [infinity] and nonlinear approaches*. Hoboken, N.J. : Wiley-Interscience, Hoboken, N.J., 2006.

[49] Smith, T. B. Electricity theft: a comparative analysis. *Energy Policy*, 32(18):2067–2076, 2004. doi: 10.1016/S0301-4215(03)00182-4.

[50] Taaffe, K., Geunes, J., and Romeijn, H. E. Target market selection and marketing effort under uncertainty: The selective newsvendor. *European Journal of Operational Research*, 189(3):987–1003, sep 2008. ISSN 0377-2217. doi: 10.1016/J.EJOR.2006.11.049. URL https://www-sciencedirect-com.ezproxy.is.ed.ac.uk/science/article/pii/S0377221707006716.

[51] Taylor, J. W. Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, 54(8):799–805, aug 2003. ISSN 0160-5682. doi: 10.1057/palgrave.jors.2601589. URL http://link.springer.com/10.1057/palgrave.jors.2601589.

[52] Taylor, J. W. Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, 204(1):139–152, 2010. ISSN 03772217. doi: 10.1016/j.ejor.2009.10.003. URL http://www.sciencedirect.com/science/article/pii/S037722170900705X.

[53] Tenenbaum, B. and Besant-Jones, J. The California power crisis : Lessons for developing countries, 2001. URL http://documents.worldbank.org/curated/en/285361468778820018/The-California-power-crisis-Lessons-for-developing-countries.

[54] The World Bank. The World Bank Enterprise Surveys. URL http://www.enterprisesurveys.org.

[55] The World Bank. Implementing power rationing in a sensible way : lessons learned and international best practices. 2005.

[56] The World Bank. Does Private Sector Participation Improve Performance in Electricity and Water Distribution? Technical report, 2008.

[57] The World Bank. Reducing technical and non-technical losses in the power sector. Technical report, The World Bank, 2009.

[58] The World Bank. The State of Electricity Access Report 2017. Technical report, 2017.

[59] Vira Chankong. and Haines, Y. Y. *Multiobjective decision making : theory and methodology*. Dover Publications, 2008. ISBN 0486462897.

[60] Wächter, A. and Biegler, L. T. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1): 25–57, mar 2006. doi: 10.1007/s10107-004-0559-y. URL http://link.springer.com/10.1007/s10107-004-0559-y.

[61] Williams, J. and Ghanadan, R. Electricity reform in developing and transition countries: A reappraisal. *Energy*, 31(6-7):815–844, may 2006. ISSN 03605442. doi: 10.1016/j.energy.2005.02.008. URL http://www.sciencedirect.com/science/article/pii/S0360544205000423.

[62] Zhu, J. Optimal Load Shedding. In *OPTIMIZATION OF POWER SYSTEM OPERATION*, pages 437–482. John Wiley & Sons, Inc, Hoboken, NJ, USA, jan 2015. doi: 10.1002/9781118887004.ch11. URL http://doi.wiley.com/10.1002/9781118887004.ch11.

[63] Zimmerman, R. D., Murillo-Sanchez, C. E., and Thomas, R. J. MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education. *IEEE Transactions on Power Systems*, 26(1):12–19, feb 2011. ISSN 0885-8950. doi: 10.1109/TPWRS.2010.2051168. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5491276.